



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>5</sup> :</b> C12Q 1/44, 1/68, C07H 21/00	<b>A1</b>	<b>(11) International Publication Number:</b> WO 93/15221 <b>(43) International Publication Date:</b> 5 August 1993 (05.08.93)
<b>(21) International Application Number:</b> PCT/US93/00977 <b>(22) International Filing Date:</b> 29 January 1993 (29.01.93) <b>(30) Priority data:</b> 07/827,208                      29 January 1992 (29.01.92)      US 07/857,059                      24 March 1992 (24.03.92)        US 07/974,409                      12 November 1992 (12.11.92)    US <b>(60) Parent Applications or Grants</b> <b>(63) Related by Continuation</b> US                                      07/974,409 (CIP) Filed on                                12 November 1992 (12.11.92) US                                      07/857,059 (CIP) Filed on                                24 March 1992 (24.03.92) US                                      07/827,208 (CIP) Filed on                                29 January 1992 (29.01.92) <b>(71) Applicants (for all designated States except US):</b> HITACHI CHEMICAL CO., LTD. [JP/JP]; 48 Wadai, Tsukuba-City, Ibaraki (JP). HITACHI CHEMICAL RESEARCH CENTER, INC. [US/US]; 1003 Health Sciences Road West, Irvine, CA 92715 (US).		<b>(72) Inventors; and</b> <b>(75) Inventors/Applicants (for US only):</b> AKITAYA, Tatsuo [JP/JP]; 20-2, Hirasawa Terasu, 16-15-2, 1-chome, Takasuzu-machi, Hitachi Ibaraki (JP). COOPER, Allan [US/US]; 3607 N.E. 172nd Avenue, Bellview, WA (US). MITSUHASHI, Masato [JP/US]; 8 Brookmont, Irvine, CA 92714 (US). <b>(74) Agents:</b> ALTMAN, Daniel, E. et al.; Knobbe, Martens, Olson and Bear, 620 Newport Center Drive, 16th Floor, Newport Beach, CA 92660 (US). <b>(81) Designated States:</b> CA, JP, KR, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). <b>Published</b> <i>With international search report.</i>
<b>(54) Title:</b> METHOD FOR MEASURING MESSENGER RNA  <b>(57) Abstract</b>  <p>The present invention provides a method for detecting and quantifying mRNA in a sample. The mRNA that can be detected has a unique sequence. The method includes immobilizing a first polynucleotide to an insoluble support. The first polynucleotide has a first sequence that hybridizes to the unique sequence on the mRNA. After immobilization of the first polynucleotide, the sample is applied to the insoluble support under conditions that allow the unique sequence on the mRNA to hybridize with the first polynucleotide. Thereafter, a labeled probe can be applied to the insoluble support. This labeled probe has a sequence thereon that hybridizes to a portion of the mRNA other than the unique sequence. The application of the labeled probe is performed under conditions that allow the labeled probe to hybridize with mRNA immobilized on said support, if present. Finally, the amount of the labeled probe immobilized on the support is measured to provide an indication of the amount of mRNA present in the sample.</p>		

# INTERNATIONAL SEARCH REPORT

International Application No. PCT/US91/08769

<b>I. CLASSIFICATION OF SUBJECT MATTER</b> (In several classification symbols apply, indicate all) *		
According to International Patent Classification (IPC) or to both National Classification and IPC		
IPC(5): C07H 17/00; C 08 G 63/48		
U.S.C1: 525/54.2; 54.3; 536/27, 28, 29		
<b>II. FIELDS SEARCHED</b>		
Minimum Documentation Searched ?		
Classification System	Classification Symbols	
U.S.C1	525/54.2; 54.3; 536/27, 28, 29	
Documentation Searched other than Minimum Documentation to the extent that such Documents are included in the Fields Searched *		
<b>III. DOCUMENTS CONSIDERED TO BE RELEVANT</b> †		
Category *	Citation of Document, †† with indication, where appropriate, of the relevant passages ‡	Relevant to Claim No. ‡
Y	US, A, 4,757,141 (FUNG et al) 12 July 1988 see columns 5 and 6 and column 3, lines 29-43.	1-63
Y	WO, A, 8600074 (TCHEN) 03 January 1986, see page 8.	1-48
Y	US, A, 3,440,190 (MELBY) 22 April 1969, see column 2, lines 1-36.	49-63
<p>* Special categories of cited documents: ††</p> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier document but published on or after the international filing date</p> <p>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p> <p>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>"X" document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step</p> <p>"Y" document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.</p> <p>"G" document member of the same patent family</p>		
<b>IV. CERTIFICATION</b>		
Date of the Actual Completion of the International Search		Date of Mailing of this International Search Report
09 January 1992		15 JAN 1992
International Searching Authority		Signature of Authorized Officer
ISA/US		James O. Wilson

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FR	France	MR	Mauritania
AU	Australia	GA	Gabon	MW	Malawi
BB	Barbados	GB	United Kingdom	NL	Netherlands
BE	Belgium	GN	Guinea	NO	Norway
BF	Burkina Faso	GR	Greece	NZ	New Zealand
BG	Bulgaria	HU	Hungary	PL	Poland
BJ	Benin	IE	Ireland	PT	Portugal
BR	Brazil	IT	Italy	RO	Romania
CA	Canada	JP	Japan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SK	Slovak Republic
CI	Côte d'Ivoire	LI	Liechtenstein	SN	Senegal
CM	Cameroon	LK	Sri Lanka	SU	Soviet Union
CS	Czechoslovakia	LU	Luxembourg	TD	Chad
CZ	Czech Republic	MC	Monaco	TG	Togo
DE	Germany	MG	Madagascar	UA	Ukraine
DK	Denmark	ML	Mali	US	United States of America
ES	Spain	MN	Mongolia	VN	Viet Nam
FI	Finland				

## METHOD FOR MEASURING MESSENGER RNA

### Background of the Invention

#### Field of the Invention

This invention is related to a method for measuring messenger RNA (mRNA) and polynucleotide-immobilized supports useful in this method. Specifically, this invention is related to measuring methods and reagents in which various types of mRNA can be measured at the same time. In one embodiment of this invention, the mRNA need not be purified from the cell, and the mRNA can be rapidly quantified with high sensitivity.

#### Description of Previous Techniques

Previously well-known techniques for measuring mRNA levels include the Northern Blot Method (See, e.g., J. Sambrook, et al., Molecular Cloning, 2nd ed., Cold spring Harbor Laboratory Press, 1989 hereafter referred to as "Molecular Cloning"), the Dot Blot Method (See, e.g., Molecular Cloning), Ribonuclease Protection Assay (See, e.g., Molecular Cloning), and the Reverse PCR Method (H.A. Erlich (ed.), PCR Technology-Principles and Applications for DNA Amplification, Stockton Press, New York, NY, 1989). The disclosures of all of these previous references are hereby incorporated by reference.

The Northern Blot Method detects mRNA which is separated by molecular weight using electrophoresis. The separated molecules are immobilized on a membrane and mRNA levels are determined by hybridization of marker DNA probes. With this method, the molecular weight of the subject mRNA must generally be known. By using this method a complex between a labeled probe, mRNA and an insoluble membrane is formed. The amount of label in the complex at the appropriate site on the membrane is then measured colorimetrically or by radioactive emissions.

The Dot Blot Method detects mRNA by hybridization with a labeled probe after purified mRNA is immobilized as a spot on a membrane. This method can be used to rapidly examine many samples at one time. However, both the Dot Blot and the Northern Blot methods require the use of a specific labeled probe for each mRNA which is desired to be detected.

The Ribonuclease Protection Assay method is used to detect mRNA after it is initially purified. The purified mRNA is hybridized with a labeled RNA probe, and then treated with a ribonuclease that specifically digests single stranded RNA. The labeled RNA probe will form a duplex in the solution and be protected from digestion. This is then immobilized on a membrane after separation of the double-stranded RNA/probe complex.



-2-

The reverse PCR method is used where cDNA is synthesized using reverse transcriptase and purified mRNA. This cDNA can be immobilized to a membrane. The synthesized cDNA is then amplified using standard PCR (Polymerase Chain Reaction) methods. After separating cDNA according to its molecular weight by electrophoresis, it becomes a single band. The DNA is then immobilized on a membrane, and detected by hybridization with a labeled DNA probe. However, after PCR, no quantitative information remains due to the amplification of the original message.

All of the foregoing methods take extended periods of time to perform since the mRNA used must be in at least a partially purified form. Also, electrophoresis of the samples and immobilization to a membrane is necessary. Both of these steps reduce the reliability of quantification of the mRNA.

Thus, there remains a need to resolve the above problems, and to provide a method of measuring for rapidly determining messenger RNA levels and a measuring reagent useful in such a method.

#### Summary of the Invention

In the mRNA assay methods of the present invention, a first and a second polynucleotide probe can be used. The inventors have discovered that immobilization of probes (a first polynucleotide) rather than samples, on a water-insoluble support can provide a basis for the rapid quantification of multiple unpurified samples containing mRNA. Applicants have provided an efficient method of identifying appropriate specific probes for this purpose. In preferred forms of the present invention, the first polynucleotide probe has a nucleotide sequence complementary to a region which is unique in the sample to the mRNA and does not bind other portions in the target mRNA. The inventors also discovered that a single kind of second polynucleotide probe can be used for simultaneous quantification of various mRNAs. However, any of a variety of probes that bind to a different region of the mRNA than the first probe can be used for this purpose. In this regard, the inventors have found that a second polynucleotide probe containing sequences complementary to a polyadenylic acid tail of mRNA can be used in such a simultaneous assay method. Based on these and other discoveries, the inventors established the present invention.

One embodiment of the present invention is a highly sensitive, quantitative and rapid method for detecting and quantifying mRNA in a sample without the need to purify mRNA from cells. In this embodiment of the invention, the method comprises the following steps:

- (a) identifying a polynucleotide sequence that is unique to the mRNA:

-3-

(b) immobilizing a first polynucleotide to an insoluble support, the first polynucleotide having a first sequence complementary to the sequence unique to the mRNA;

(c) incubating the sample with the insoluble support under conditions wherein the unique sequence will hybridize with the first polynucleotide, thereby immobilizing mRNA present in the sample to the insoluble support;

(d) washing non-immobilized components of the sample from the insoluble support;

(e) labeling mRNA on the support in a manner that label is incorporated onto the support related to the amount of mRNA on the support; and

(f) measuring the amount of label immobilized on the support.

In a preferred embodiment, the above step of labeling mRNA on the support involves incubating a labeled probe, most preferably poly-d(T), with the insoluble support under conditions wherein the labeled probe will hybridize with mRNA on the support, and the labeled probe bears a label and is complementary to a part of the mRNA other than a part of the mRNA complementary to the first probe, thereby immobilizing to the insoluble support the label on the labeled probe which has hybridized with the mRNA; and thereafter washing any non-immobilized labeled probe from the insoluble support.

In addition, the above method of labeling nucleic acids in step (e) can preferably involve labeling the oligonucleotides with a nucleic acid stain, most preferably from the group consisting of ethidium bromide, yoyo-1 and toto-1.

In a further preferred embodiment, the above method can utilize a label, wherein the label comprises a radionuclide, or alternatively biotin. In the method using biotin, the incubation can preferably be with enzyme-conjugated streptavidin; followed by measurement of the amount of enzyme bound to the insoluble support. In this embodiment step (f) can additionally comprise adding streptavidin bound to alkaline phosphatase to the insoluble support.

In still another embodiment of the preferred method, the mRNA can code for a  $\beta$  receptor, and wherein the first polynucleotide comprises the sequence of SEQ ID NO:6. An alternative embodiment of the above method includes mRNA coding for the  $\alpha$  subunit of Gs protein, wherein the first polynucleotide comprises a sequence complementary to a sequence selected from the group consisting of SEQ ID NO:1, 145-146, 209-212, 268-287, 705 and 708.

Yet another alternative mRNA sequence that can be used with the present invention method codes for the Gi-2 protein, wherein the first polynucleotide comprises a sequence complementary to a sequence selected from the group consisting of SEQ ID NO:703, 806, 142, 143, 187-197 and 238-253.

5 The present method can also preferably involve the use mRNA coding for substance P receptor, wherein the first polynucleotide comprises the sequence of SEQ ID NO:5.

mRNA encoding Gi-3 protein can also be used in the present method, wherein the first polynucleotide comprises a sequence complementary to a sequence selected from the group consisting of SEQ ID NO:704, 707, 144, 198-208 and 254-267.

10 Still another embodiment of the above method can involve the mRNA coding for the Gi-1 protein, wherein the first polynucleotide comprises a sequence complementary to a sequence selected from the group consisting of SEQ ID NO:702, 805, 141, and 224-237. An even further preferred embodiment includes mRNA coding for G protein, wherein the first polynucleotide comprises a sequence complementary to a sequence selected from the group consisting of SEQ ID NO:700, 701, 139, 140, 148-159, 162-172, 176-186 and 3-4.

15 Most preferably, the first sequence hybridizes to a unique sequence of the mRNA and not to other polynucleotides present in the sample, and even more preferably, the sample comprises a cell lysate.

In an alternative preferred embodiment, the label in the above method can be measured by light emitted therefrom, and step (f) of the method comprises measuring the amount of light emitted by the label. The method of measuring the light can preferably involve: recording the amount of light on film; and measuring the exposure of this film using a densitometer. In addition, the light measurement can alternatively comprise adding ATTOPHOS and measuring fluorescence emitted using a fluorimeter.

25 The present invention method in step (a) above can advantageously include a computer program, with the computer program preferably involving use of an H-site model. The H-site model of step (a) is preferably used to identify the first polynucleotide probe of the method by the steps of:

specifying a minimum melting temperature for the first nucleotide probe and the nucleotide sequence specific to the organism;

30 specifying a nucleation threshold that places a minimum value on the number of base pairs at any nucleation site;

-5-

determining the melting temperatures ( $T_m$ ) of the first nucleotide probe and the sequence specific to the organism at every possible hybridization point; and

selecting the nucleotide probe having the highest  $T_m$  value.

5 The computerized method of identifying the first polynucleotide probe can preferably have the melting temperature determined by the formula:

6  $T_m = 81.5 - 16.6(\log[Na]) - 0.63\%(\text{formamide}) + 0.41(\%(G+C)) - 600/N$ , wherein  $\log[Na]$  is the log function of the sodium concentration,  $0.063\%(\text{formamide})$  is the concentration of formamide,  $\%(G+C)$  is the percentage of matched GC base pairs, and N is the probe  
10 length.

Additionally the method of step (a) can be used to identify the labeled probe referred to above using the H-site computerized model by the steps of:

specifying a minimum melting temperature for the labeled probe and the nucleotide sequence specific to the organism:

15 specifying a nucleation threshold that places a minimum value on the number of base pairs at any nucleation site:

determining the melting temperatures ( $T_m$ ) of the first nucleotide probe and the sequence specific to the organism at every possible hybridization point; and

20 selecting the nucleotide probe of the proper length having the lowest  $T_m$  value.

A method of identifying the second nucleotide probe can involve determining the melting temperature by the formula:

25  $T_m = 81.5 - 16.6(\log[Na]) - 0.63\%(\text{formamide}) + 0.41(\%(G+C)) - 600/N$ , wherein  $\log[Na]$  is the log function of the sodium concentration,  $0.063\%(\text{formamide})$  is the concentration of formamide,  $\%(G+C)$  is the percentage of matched GC base pairs, and N is the probe length.

30 A different embodiment of the present invention is a method of inhibiting RNase activity in a sample containing RNA in a buffer system containing SDS and EDTA, comprising adding vanadyl ribonucleoside complex to the sample.

Preferably, there is at least a five-fold molar excess of EDTA over RNA in the sample and most preferably, the vanadyl ribonucleoside complex is added in combination with proteinase K. In an alternative preferred embodiment, the vanadyl ribonucleoside complex is added in combination with RNasin.

Yet another different embodiment of the present invention is a polynucleotide-immobilized support that is useful in detecting and quantifying the amount of  $\beta$  receptor mRNA in a sample, comprising:

- an insoluble support,
- 5 a first polynucleotide bound to the support, the first polynucleotide comprising a sequence that is complementary to a sequence unique to  $\beta$  receptor mRNA.

More preferably, the sequence unique to  $\beta$  receptor mRNA comprises a sequence selected from the group consisting of SEQ ID NO:6.

- 10 Still another preferred embodiment of the present invention is a polynucleotide-immobilized support that is useful in detecting and quantifying the amount of jun oncogene mRNA in a sample, comprising:

- an insoluble support;
- 15 a first polynucleotide bound to the support, the first polynucleotide comprising a sequence that is complementary to a sequence unique to jun oncogene mRNA.

Preferably, the sequence unique to the jun oncogene mRNA comprises a sequence selected from the group consisting of SEQ ID NO:11-35, 39-53, 62-138, 310-342, 600-606, 614-620, and 628-634.

- 20 Another embodiment of the present invention is a polynucleotide-immobilized support that is useful in detecting and quantifying the amount of G protein mRNA in a sample, comprising:

- an insoluble support;
- 25 a first polynucleotide bound to the support, the first polynucleotide comprising a sequence that is complementary to a sequence unique to G protein mRNA.

Preferably, the sequence unique to G protein mRNA comprises a sequence selected from the group consisting of 1, 2, 700-709, 805, 806, 139-159, 162-172 and 176-309.

- 30 Another preferred embodiment of the present invention is a polynucleotide-immobilized support that is useful in detecting and quantifying the amount of Substance P receptor mRNA in a sample, comprising:

- an insoluble support:

a first polynucleotide bound to the support, the first polynucleotide comprising a sequence that is complementary to a sequence unique to Substance P receptor mRNA.

Preferably the sequence unique to Substance P receptor mRNA comprises the sequence of  
5 SEQ ID NO:5.

#### Brief Description of the Figures

Figure 1 illustrates an immobilization procedure of a first polynucleotide probe onto an insoluble support by the maleimide method.

Figure 2 illustrates an immobilization procedure of a first nucleotide probe onto an  
10 insoluble support by the carbodiimide method.

Figure 3 schematically shows the principle of an mRNA quantification method in accordance with the present invention.

Figure 4 (I) shows a photographic image of an X-ray film exposed to chemiluminescent light derived from the human  $\beta_2$  adrenergic receptor cDNA probe on a  
15 Northern membrane, in which various concentrations of the human  $\beta_2$  adrenergic receptor mRNA were immobilized. (II.A) shows a photographic image of an X-ray film exposed to chemiluminescent light derived from the oligo-(dT) probe in accordance with the present invention in which various concentrations of the human  $\beta_2$  adrenergic receptor mRNA were trapped onto the first oligonucleotide-immobilized plastic plate. (II.B) shows a densitogram  
20 of the image shown in (II.A.). (II.C) graphically depicts the relationship between relative intensity of chemiluminescent signals and concentrations of mRNA of the image shown in (II.A).

Figure 5 graphically depicts the linear relationship between relative intensity of chemiluminescent signals and concentrations of mRNA in a separate experiment from that  
25 shown in Figure 4.II.

Figure 6 graphically depicts a calibration line for the rat Gs protein mRNA of the invented method.

Figure 7 shows a photographic image of a polaroid film exposed to chemiluminescent light derived from oligo-(dT) probe by the invented method, in which Gi-2 and GS protein  
30 mRNA in various cells were trapped onto the first polynucleotide-immobilized plastic plate.

Figure 8 shows a photographic image of an X-ray film exposed to chemiluminescent light derived from the Gi-2 and Gs protein cDNA probes on the Southern membrane, in which PCR-amplified G protein specific DNA from 4 different cell lines was immobilized.

Figure 9 graphically depicts a calibration line for the human substance P receptor mRNA of the invented method.

Figure 10 shows a gel showing the effect of various RNase inhibitors on mRNA preparations containing SDS and EDTA.

5        Figure 11 shows a gel of five different G protein oligonucleotides amplified with the  $G_2$  and  $G_4$  PCR primers and also provides Southern blots using each of the five G protein sequences as a probe.

10        Figure 12 shows a gel of samples containing varying amounts of each of five different G protein oligonucleotides (as indicated by number of "+" symbols) amplified with the  $G_2$  and  $G_4$  PCR primers and also provides Southern blots using each of the five G protein sequences as a probe.

Figure 13 shows a gel of  $\lambda$ ZAP cDNA libraries from rat pituitary (P), kidney (K) and intestinal (I) amplified with each of five different G protein primers, as indicated.

15        Figure 14 shows a gel of 500 bp DNA from cDNAs of human IM9 and Jurkat cells amplified with  $G_2$  and  $G_4$  PCR primers.

Figure 15 is a graph showing the relationship between YOYO-1 fluorescence and amount of immobilized oligonucleotide.

Figure 16 is a graph showing the time course of YOYO-1 analysis.

Figure 17 is a graph showing the dose response of YOYO-1 concentration.

20        Figure 18 is a bar graph showing the reproducibility of YOYO-1 staining based on quantitation of immobilized oligonucleotides.

Figure 19 exhibits a Polaroid picture of agarose gel electrophoresis of the PCR amplified ds-cDNA of jun oncogene from human peripheral blood leukocytes.

25        Figure 20 is a simplified block diagram of a computer system illustrating the overall design of this invention;

Figures 21A-21C show display screen representations of the main oligoprobe design station dialog windows of this invention;

Figure 22 is a flow chart of the overall invention illustrating the program and the invention's sequence and structure:

30        Figure 23 is a display screen representation of the Mitsuhashi probe selection diagram:

Figure 24 is a display screen representation of the probeinfo and matchinfo window;

Figure 25 is a display screen representation of the probesedit window;

Figure 26A is a printout of the probesedit output file:

Figure 27 is a flow chart of the overall k\_diff program of the Mismatch Model of this invention, including its sequence and structure;

Figure 28 is a flow chart of the k\_diff module of this invention;

Figure 29 is a flow chart of the hashing module of this invention;

Figure 30 is a flow chart of the tran module of this invention;

Figure 31 is a flow chart of the let\_dig module of this invention;

Figure 32 is a flow chart of the update module of this invention;

Figure 33 is a flow chart of the assembly module of this invention;

Figure 34 is a flow chart of the seqload module of this invention;

Figure 35 is a flow chart of the read1 module of this invention;

Figure 36 is a flow chart of the dig\_let module of this invention;

Figure 37 is a flow chart of the q\_colour module of this invention;

Figure 38 is a flow chart of the hit\_ext module of this invention;

Figure 39 is a flow chart of the colour module of this invention;

Figure 40 is the first page of a printout of a sample file containing the output of the Mismatch Model program of this invention;

Figure 41 is a flow chart of the H-Site Model, stage I, covering the creation of a preprocessed preparation file of this invention;

Figure 42 is a flow chart of the H-Site Model, stage II, covering the preparation of the target sequence(s);

Figure 43 is a flow chart of the H-Site Model, stage III, covering the calculation of MPSD data;

Figure 44A is the first page of a printout of a sample file containing output of Mismatch Model program;

Figure 44B is the first page of a printout of a sample file containing output of H-Site Model program;

Figure 45 is a flow chart of the processing used to create the Mitsuhashi probe selection diagram (MPSD);

Figure 46 is a flow chart of processing used to create the matchinfo window;

Figure 47 is a printout of a sample target species file;

Figure 48 is a printout of a sample preparation file.

#### Detailed Description of the Invention

##### Target mRNA



-10-

Previously, in order to genetically diagnose an infectious disease, procedures utilizing DNA were generally performed. Although, with DNA diagnosis, the existence of an external gene can be confirmed, it cannot be determined if the known gene is in an unexpressed carrier state or expressed in the diseased state.

5 In living organisms, once a gene is activated, the DNA coding for that gene is transcribed into mRNA, and the mRNA is subsequently translated into protein. Thus, the activation of the external gene can be determined by measuring the levels of mRNA.

Currently, to genetically diagnose a protein deficiency state, most experimenters have limited themselves to analysis at the DNA level. However, this method cannot be used to  
10 determine abnormalities of transcription or translation. If mRNA levels could be measured easily, the diseased state might be more precisely recognized. In a diseased state where mRNA cannot be detected, it can be said that there is a defect in transcription, while if normal levels of mRNA are detected, there is probably an abnormality in translation.

In addition, the amount of mRNA produced in the cell reflects the amount of  
15 expression of a gene, so it would be useful in the course of an abnormal condition, such as an illness or drug reaction, to examine mRNA levels. As discussed in greater detail below, measuring any of a large number of mRNAs can provide clinically useful information. Particular examples include mRNA for the  $\beta$  receptor, substance P, G protein and oncogenes, such as jun.

20 For the aforementioned reasons, the ability to measure mRNA would be useful for diagnosing and recognizing the pathophysiology of various disease states. This method would be especially useful for diagnosis of hereditary diseases, cancer, and infectious diseases.

A preferred target of this invention is mRNA of eukaryote cells which contain  
25 polyadenylic acid tail on their 3'-terminal end. Once the specific nucleotide sequence of the mRNA is known, virtually any mRNA can be measured using the methods of the present invention. Examples of such mRNAs are human G protein, human  $\beta_2$  receptor, substance P, lymphocyte surface antigen, immunoglobulin, collagen, and adrenaline receptor.

#### Test Samples

30 Test samples containing target mRNA can be obtained from a variety of sources. Preferred sources include living samples or cell solutions which have been treated to inactivate ribonuclease (RNase) activity. Purification of mRNA for standard samples can be conducted according to the methods usually associated with Northern Blot procedures

or Dot Blot procedures method. Kits for the purification of mRNA are commercially available.

5 A test sample can be prepared from a living sample by treatment with 10 mM of ethylenediaminetetraacetic acid (EDTA) (pH 8.0), 0.2 M NaCl, 0.5% of sodium dodecyl sulfate (SDS), 500 Unit/ml of RNase inhibitor, 10mM of Vanadyl Ribonucleoside Complex  
10 and 200 µg/l of Proteinase K (hereafter called Lysis buffer) to lyse the cells and inhibit RNase activity. After lysis of the cells, NaCl concentration is adjusted to 0.5M.

In the methods of the present invention that involve probing samples containing RNA, or in any procedure in which it is desired to prevent the degradation of RNA, it is  
10 advantageous to inhibit the activity of any ribonucleases (RNases) which may be present. One Rnase inhibitor which can be used is Vanadyl Ribonucleoside Complex (VRC) (Bethesda Research Laboratories, Gaithersburg, MD). VRC has been reported to be useful during cell fractionation and in the preparation of RNA, and has been shown not to interfere with the phenol extraction or ethanol precipitation of RNA. In addition, VRC  
15 does not affect other cytoplasmic components of cells. Therefore, VRC is an ideal inhibitor of RNase in many experimental procedures.

However, prior art procedures for inhibiting RNases with VRC taught that VRC should not be used in buffer systems containing EDTA or SDS, which are commonly used in the field of molecular biology. The reason for this prohibition was that it was believed  
20 that the Complex would dissociate in the presence of these buffers, leading to a loss of RNase inhibiting activity. In fact, the BRL insert accompanying the VRC product recommends that a five- to ten-fold molar excess of EDTA be added to an RNA solution containing VRC in order "destroy" the VRC prior to ethanol extraction of the RNA from the solution. The apparent inability to use VRC together with common buffers that include  
25 EDTA and SDS thus presented a major impediment to the exploitation of VRC as an RNase inhibitor.

We have discovered, however, that VRC is in fact an effective RNase inhibitor even in the presence of SDS and/or EDTA. Thus, VRC can be used in assays which use buffers including SDS or EDTA, where heretofore it was believed that VRC would not be effective  
30 in such systems. VRC is an effective RNase inhibitor at the concentrations of EDTA and SDS that are normally used when manipulating RNA or when performing a variety of other molecular biology techniques. For example, we have found that VRC effectively inhibits RNase in a buffer solution comprising approximately 1 mM EDTA. We also found VRC to be effective in .5% SDS solutions, and is believed to be effective in solutions ranging up

-12-

to 2% SDS, more preferably up to 1% SDS. VRC can, of course, also be used in other solutions including EDTA and SDS.

We have found that VRC is a particularly potent inhibitor of RNases when used in combination with Proteinase K. Proteinase K is also available from BRL. As shown in the gel in Figure 10, mRNA prepared from U937 cells (human macrophage cell line) was protected from RNase degradation by a combination of VRC and Proteinase K. Lane 1 of the gel shows the mRNA from a cell preparation which included VRC, Proteinase K, and RNasin, while lane 2 represents the mRNA from a cell preparation that included VRC and Proteinase K. RNasin is available from Promega of Madison, WI. The distinct band 10 in lanes 1 and 2 matches the band seen in lane 7, which contains pure clonal cDNA from an RNase-free preparation of U937 cells, thus showing that the mRNA in the preparations of lanes 1 and 2 did not experience substantial mRNA degradation.

A comparison of lanes 1 and 2 with lanes 3-6 shows that VRC in combination with Proteinase K inhibits RNase activity in the above-mentioned mRNA preparation from U937 cells to a far greater extent than either Protein K alone (lane 4), Proteinase K in combination with RNasin (lane 3), or a commercial RNase-inhibiting preparation sold under the name "FastTrack" (available from In Vitrogen of San Diego, CA) (lane 5). None of lanes 3-5 exhibit the distinct band 10 representing undegraded mRNA. On the contrary, lane 4 (Proteinase K alone) and lane 5 (FastTrack) have the same smear of degraded mRNA as band 20 in lane 6 (no inhibitors). The gel shown in Figure 10 also shows the effectiveness of VRC in a buffer solution of 1 mM EDTA and .5% SDS, which is the buffer used in the U937 cell preparations tested, since lane 2 (VRC and Proteinase K) shows less mRNA degradation than lane 4 (Proteinase K alone) or lane 3 (Proteinase K and RNasin).

#### **First Polynucleotide Probe**

In the practice of the present invention, a first polynucleotide probe is bound to an insoluble support. This first polynucleotide is used to trap the target mRNA to a solid support. Thus, the first polynucleotide probe comprises a nucleotide sequences which can hybridize with a sequence of the mRNA that does not occur in other RNA in the sample. Such a sequence is referred to herein as a unique region of the mRNA. Preferably, the first polynucleotide probe does not bind to portions of the target mRNA other than the unique region.

A unique region can be identified by any of several methods. These include: (1) comparing the nucleotide sequence between several potential mRNAs and their target mRNAs; (2) getting a candidate sequence which can be compared with other nucleotide

sequences; and (3) computer searching for nucleotide sequences which do not pair with the target mRNA. Information on various known mRNA sequences can be obtained from several genetic information databases.

5 The first polynucleotide probe is preferably an oligodeoxyribonucleotide (DNA) rather than an oligoribonucleotide (RNA) since DNA is more stable than RNA. The number of nucleotides is not restricted; however, if an oligodeoxynucleotide is used as the first probe, a preferred length is 17-40 nucleotides. If a cDNA of target RNA is available, a part or the entire cDNA can be used as the first nucleotide probe. The first nucleotide probe can also be easily manufactured by using a DNA synthesizer or by a commercial  
10 manufacturer.

In this invention, the first nucleotide probe used is immobilized to an insoluble support. It may be chemically modified for immobilization, such as by adding an amino group onto the 5'-terminal or 3'-terminal of the polynucleotide.

#### Insoluble Support

15 In the practice of the present invention, the first nucleotide probe is immobilized to an insoluble support. The insoluble support can be any of a variety of insoluble materials, such as a plastic plate, membrane filter, microtiter plate or any other insoluble material to which polynucleotides can be attached. The insoluble support is preferably made of a material which can immobilize the first nucleotide probe by covalent bonds. For example,  
20 a plastic plate, such as polystyrene, which has a carboxyl group or amino group on the surface thereof can be used. Such plates are available commercially. However, procedures for introducing these residues are also well known in the art. One commercially available plastic plate which has carboxyl group, on its surface is Sumilon microtiter plate MS-3796F made by Sumitomo Bakelite. A plastic plate having an amino group on the surface is the  
25 Sumilon microtiter plate MS-3696F.

#### Immobilization of First Polynucleotide

In the practice of the present invention, the first polynucleotide probe is immobilized onto the insoluble support. Various methods of immobilizing polynucleotides to the insoluble support are available, including covalent binding, ionic binding, and the physical  
30 absorbance method. However, the covalent binding method is preferred. Thus, in certain embodiments of the present invention, the polynucleotides are immobilized to microtiter plates which exhibit functional groups, such as carboxyl residues, amine residues, or hydroxyl residues on the surface thereof.

In a preferred procedure for immobilization of the polynucleotide to an insoluble support exhibiting a functional group, the 5'-terminal end of the polynucleotide is covalently linked to the functional group. Any of a variety of methods for covalent binding of polynucleotides to these functional groups can be used. Examples of preferred well-known methods include the maleimide method and carbodiimide method.

The maleimide method, illustrated in Figure 1, involves the reaction between a substance containing a maleimide group and another material containing a sulfhydryl residue (SH). In order to attach the 5' end of a polynucleotide to an immobilized support using the maleimide method, the 5' end of the polynucleotide is reacted with a maleimide compound. A suitable maleimide compound is sulfosuccinimidyl-4-(N-maleimidomethyl) cyclohexane-1-carboxylate (sulfo-SMCC).

The SH residue is provided on the support by a reaction between a support having an amino group and succinimidyl-S-acetylthioacetate (SATA), followed by deacetylation using hydroxylamine ( $\text{NH}_2\text{OH}$ ). (Sulfo-SMCC and SATA are readily available from a variety of commercial sources, including the Pierce Company.) The resulting SH group on the support can then react with the maleimide group on the 5' end of the polynucleotide to form a polynucleotide-immobilized support. One problem we have experienced in the use of the maleimide method is that the SH group on the plate can react not only with an amino group at the 5' end of the polynucleotide, but can also react with primary amino groups on the purine bases, adenine and guanine. In order to assure that the polynucleotides are immobilized at their 5' ends, the amino groups on the purine bases can be protected by pairing the polynucleotide to a complementary polynucleotide prior to immobilization. After immobilization, the complementary polynucleotide can be removed through denaturation, such as through heating.

The carbodiimide method is illustrated in Figure 2. This method involves a reaction between an amino group and a material containing a carboxyl residue using a carbodiimide compound. An example of a carbodiimide compound is 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (hereafter called EDC). This reaction can be enhanced with N-hydroxysulfosuccinimide (hereafter called Sulfo-NHS). Both EDC and Sulfo-NHS are available from well known commercial sources, including the Pierce Company.

In the practice of a preferred carbodiimide method for attaching polynucleotides to a support, a support having a carboxyl residue attached is used. EDC is activated by reaction with Sulfo-NHS. This activated EDC is reacted with support containing surface-

-15-

bound carboxyl residues. This can then be reacted with polynucleotides having an amino group at their 5'-terminal ends, resulting in a polynucleotide-immobilized support.

In order to assure that the polynucleotides are immobilized at their 5' end, primary amino groups on adenyl, guanyl and cytosyl groups can be protected by hybridizing the nucleotide to a complementary polynucleotide prior to immobilization. After immobilization, the complementary polynucleotide can be removed through denaturation, such as through heating.

We have found that non-specific binding of activated amino or carboxyl residues on insoluble supports can be effectively reduced or eliminated by treating the plates to which polynucleotides have been immobilized with a primary amine compound, preferably glycine.

#### Second Polynucleotide Probe

The second nucleotide probe contains sequences which can also hybridize to target mRNA that is trapped by the first bound nucleotide sequence. Preferably, the second polynucleotide probe hybridizes to a different portion of the target mRNA than does the first polynucleotide probe. In an especially preferred embodiment, the second nucleotide probe consists of nucleotides which are complementary to the polyadenylic acid tail of the target mRNA. The second nucleotide probe is also preferably DNA rather than RNA due to the higher stability of DNA. The number of nucleotides is not limited; however, a polynucleotide of 15-40 nucleotides is preferable. Thus a poly (dT) sequence of 15-40 nucleotides is a preferred second polynucleotide probe. Such a second polynucleotide probe would serve as a universal second probe for virtually any target mRNA from eukaryotic cells. However, if a cDNA to the target RNA is available, a part or all of this cDNA can also function as a second polynucleotide probe. The second nucleotide probe can also be easily manufactured by a DNA synthesizer or commercial manufacturer.

The second polynucleotide probe is preferably labeled in order to easily detect its presence. A variety of chemical substances are available which can label the target substance. For example, a variety of radionuclides can be used, such as the radioisotopes  $^{32}\text{P}$ ,  $^{35}\text{S}$ ,  $^3\text{H}$ , and  $^{125}\text{I}$ . Enzymes can also be used to label the second polynucleotide probes. Suitable enzymes include alkaline phosphatase, luciferase, and peroxidase. Other labels might be chemical compounds such as biotin, avidin, streptavidin, and digoxigenin. Labels which provide a colorimetric indication or a radionuclide are preferred. In a preferred embodiment, biotin is attached to the nucleotide probe, followed by an avidin, such as streptavidin, conjugated to an enzyme such as alkaline phosphatase. The presence of the

-16-

enzyme can be detected by various substrates, such as ATTOPHOS which provides a fluorescent marker that can be detected through fluorimetry.

5 In certain embodiments, part or all of the labeling compound is substituted with part or all of the second nucleotide probe. Techniques for substituting portions of polynucleotides are well known for a variety of labels and kits are commercially available for many of these. For example, one can use a DNA 3'-end Labeling Kit or oligo-deoxythymidylic acid to label the probe.

10 Nucleic acid can also be "labeled" by staining with a nucleic acid stain. Thus, where a relatively large amount of nucleic acid is present, ethidium bromide (EtBr) can be used to identify the presence of nucleic acid. This is useful within the context of the present invention, for example, if mRNA has hybridized to a short immobilized anti-sense probe then substantially more nucleic acid will be immobilized than if no mRNA had hybridized. However, more sensitive stains are more preferable. Such more sensitive stains include the various cyanine nucleic acid stains, such as POPO, BOBO, YOYO and TOTO available from Molecular Probes (California). These stains are described, e.g., in *Science*, 257:885 (1992). Particularly preferred stains for use in the context of the present invention include the shorter wavelength forms, TOTO-1 and YOYO-1, still more preferably YOYO-1. As little as four picograms of stained DNA can be detected by visible fluorescence upon stimulation with a transilluminator or hand-held UV lamp. Thus, these stains provide a particularly easy and sensitive method of identifying the presence of nucleic acids.

20 We tested the ability of YOYO-1 to stain oligonucleotides immobilized to wells as discussed herein. We first immobilized various known amounts of oligonucleotides to wells and washed to remove non-immobilized oligonucleotides. We then added 1:1000 dilution of YOYO-1 in water. We then used a fluorimeter directly without washing. The relation between pmoles of oligonucleotide is shown in Figure 15 with open circles. We also incubated the TOTO-1 stained immobilized oligonucleotides for ten minutes, washed and added water, followed by use of the fluorimeter. The washed results are shown in dark circles in Figure 15. It can be seen that washing does not significantly effect the amount of staining and that the amount of staining is related to the amount of oligonucleotide.

30 We also tested the time course of YOYO-1 staining over the course of one hour. We included a control in which no oligonucleotide was immobilized to the plate. The oligonucleotide-immobilized plate is shown in open circles in Figure 16 and the control in dark circles in that Figure. It can be seen from Figure 16 that after an initial spike, relatively constant staining is found.

We further tested the dose response of constant amounts of oligonucleotides with oligonucleotides immobilized in wells (open circles in Figure 17) and control wells with no oligonucleotide (closed circles in Figure 17). The difference between the immobilized and non-immobilized is shown in Figure 17 as triangles. It can be seen that a sharp increase in fluorescence occurs between  $10^{-4}$  and  $10^{-3}$  dilution.

We also tested the reproducibility of staining of both oligonucleotide-immobilized wells and non-immobilized wells. We repeated the experiment five times and graphically depicted the data for both oligonucleotide immobilized (+) and non-immobilized (-) wells in Figure 18. It can be seen that the data was substantially similar for each experiment.

Thus, the data depicted in Figures 15-18 shows that use of YOYO-1 as a staining agent provides a reliable indication of the amount of oligonucleotide present.

#### Application of Sample

In the preferred methods of the present invention, a sample containing mRNA is applied to the insoluble support to which the first polynucleotide probe has been bound. Any target mRNA found in the sample is allowed to hybridize with the first polynucleotide probe. This can be accomplished by incubation at temperature dependent on a variety of factors, as is well known to those with ordinary skill in the art. These factors include the length of complementary nucleotide sequences, the ratio of Guanine or Cytosine (GC contents) within the entire base content in the complementary nucleotide sequences, NaCl concentration in buffer, the number of bases which mismatch in the complementary nucleotide sequence, and the type of nucleotide. In a preferred form of this invention, the following equation can be used to calculate the preferred incubation temperature:

$$T_{inc.} = 16.6 \times \log (M) + 0.41 (GC) + 81.5 - 675/n - 15 (^\circ C).$$

In the equation shown above, M is the NaCl concentration (M) in solution. GC represents GC contents (%), n represents the length of nucleotide sequences (the number of nucleotides).

The incubation time can also be determined according to methods described in the Molecular Cloning manual.

The time for incubation is preferably from 1 hour to overnight, and the sample should preferably be gently swung during incubation. Incubation is preferably performed in an appropriate buffer solution. The same buffer used to hybridize RNA and DNA in the Northern Blot or the Dot Blot Methods can be used. The buffer is preferably prepared in a way so as not to contaminate it with RNase. If any RNase contamination is present, the activation thereof should be controlled to be as low as possible. RNase-free buffers are



commercially available, such as within mRNA purification kits or lysis buffer as shown before.

In order to eliminate RNase activity from water used in the methods of this invention, the water is preferably treated with Diethylpyrocarbonate (DEPC). The preferred  
5 DEPC treatment involves addition of 0.1% DEPC to the water, followed by storage overnight at 37°C and sterilization in an autoclave.

#### Washing Procedures

Preferably, after incubation, unbound components of the sample are washed from the insoluble support. Appropriate solutions for washing include the lysis buffer as shown  
10 before or the buffer included in the mRNA purification kit.

#### Application of Second Polynucleotide Probe

The second polynucleotide probe is applied to the insoluble support after any washing of unbound components of the sample, under conditions as described above in connection with the hybridization of mRNA in the sample with the first nucleotide probe.  
15 This is preferably followed by washing, as described above.

#### Quantitation

The amount of mRNA in the sample is quantitated by measuring the amount of second polynucleotide probe bound to the insoluble support after application thereto in accordance with the present invention. In this regard, a physical or chemical quantity or  
20 activity of the label on the second polynucleotide can be measured by several techniques, including optical density, emitted-light intensity, or radiation. The label itself can provide this indication or can require other compounds which bind thereto or catalyze the label. Other mechanisms for detecting label include the use of compounds that can chemically react against the label, detection of colored label, detection of light emission, detection of  
25 radiation or of catalytic ability.

In a preferred measurement technique, the label on the second polynucleotide is biotin. The presence of this label can be detected by reaction with the enzymes Peroxidase or Alkaline phosphatase. These enzymes can be specifically directed to biotin by conjugation with avidin or streptavidin. The presence of the enzymes is then detected by  
30 addition of an appropriate substrate to provide a detectable color-developing or light-emitting reaction. The alkaline phosphatase labeled streptavidin can be readily obtained from the commercial market. A preferred substrate for alkaline phosphatase is adamantyl-1,2-dioxetane phosphate (AMPPD). Upon reaction with the alkaline phosphatase, AMPPD

will emit light at a wavelength of 447 nm. This light can be detected in accordance with techniques known in the art.

In the reaction between alkaline phosphatase and AMPPD, an enhancer such as 5-N-tetradecanoyl-amino-fluorescein can be added. 5-N-tetradecanoyl-amino-fluorescein has the ability to convert light of 477 nm wavelength to 530 nm wavelength which is more readily detectable.

Other preferred labels include an antigen, such as digoxigenin or an antibody. An antigen can be detected by its ability to bind to an antibody directed thereto. Such antibodies, or an antibody directly used to label the second polynucleotide probe, can be detected by their ability to bind protein A. The antibody itself can be labeled directly with a radionuclide, such as  $^{125}\text{I}$ , or can be labeled by binding thereto of protein A labeled with the radionuclide. The radionuclide can then be detected in accordance with techniques well known in the art.

Well-known techniques for the detection of label include the detection of label which appears by a color-developing reaction by a spectrophotometer. Other such techniques include the detection of a light emitting reaction using X-ray film or an instant camera, and the detection of raditation using a radiation counter.

The emission reactions are recorded by X-ray film or instant camera film in the dark room. The X-ray film which is exposed by emission reactions are recorded as a blot, so this shading of the blot can be measured by a densitometer. If one uses an instant camera such as a Polaroid, the picture is read by a scanner to decide the location of the blot on the computer, measuring the shading of the blot using graphic analysis software.

To measure the amount of target mRNA in an unknown sample solution, an arithmetic plot for showing the relationship between a standard solution of target mRNA and relative concentration in a blot can be used. The standard solution for target mRNA need not be purified target mRNA. Thus, the standards can also be a mixture in which the amount of target mRNA is known.

#### Schematic Example

A schematic representation of an exemplary method in accordance with the present invention is shown in Figure 3. The target mRNA 1 is shown having a polyadenylic acid tail 2. The first nucleotide probe 3 is immobilized to an insoluble support 4, then mixed the target mRNA 1. The target mRNA 1 hybridizes with the first nucleotide probe 3, thereby immobilizing the target mRNA to the insoluble support. Then, a second nucleotide probe 5 which bears a label 6 hybridizes with a portion of the target mRNA other than the portion

-20-

hybridized with the first nucleotide probe. As shown, the second nucleotide probe 5 comprises poly (dT) and hybridizes with the polyadenylic acid tail 2 of the target mRNA 1. The resulting complex 7 can then be identified by detecting the presence of the label 6.

**Time-Savings of Invention**

5           Advantageously, the present invention provides a method of measuring mRNA which can be performed in far less time than previously known methods. Table 1 hereinbelow compares the time needed to measure DNA using the method of this invention and a Northern blot method.

TABLE 1  
Comparison For Measurement Of mRNA

	Present Invention		Northern Blot Method	
		Time (hours)		Time (hours)
5	- Lysis of Cells	1	- Purification of mRNA	48
10	- Hybridization of the first nucleotide probe to mRNA	1	- Separation of mRNA by electrophoresis with Agarose Gel	1
15			- denaturing of mRNA moving the mRNA to membrane	1
			- immobilization of mRNA to membrane	2
	- hybridization of the second nucleotide probe to mRNA	1	- hybridization of the mRNA and probe	12
20	- washing and blocking	0.2	- washing and blocking	4
	- detection	1	- detection	1
25	Total	about 5 hours		about 70 hours

As shown in Table 1, the methods of this invention generally take 5 hours or less to complete, while the Northern blot method generally requires about 70 hours. These times do not include the labeling time (about 12 hours) of the nucleotide probes nor the time for immobilization of the first nucleotide probe because these constructs can be prepared prior to the measurement.

Particular examples of target mRNA that can be measured to obtain clinically useful information are provided below. The examples included herein are intended to illustrate certain preferred embodiments of the present invention. These examples are not intended to limit the invention in any manner.

### 35 $\beta$ Receptor

A potential target for using mRNA measurement to determine a diseased state utilizes the mRNA for the  $\beta$  receptor. The  $\beta$  receptor is a protein located in human nerve tissue. In particular, abnormalities in the  $\beta_2$  receptor has been found to be closely correlated with asthma. Thus, measuring the mRNA for  $\beta_2$  receptor can be used to determine the pathophysiology of asthma patients, and could also be used to assess the effectiveness of anti-asthma agents.

**EXAMPLE 1**Measurement of Human  $\beta_2$  Receptor mRNA

## (Preparation of Calibration Line)

(1) Preparation of the First Nucleotide Probe

5 An oligodeoxyribonucleotide probe containing the nucleotide sequence shown below with an amino group attached to its 5' end was synthesized by Genosys, Inc. (TX). In this sequence, A, C, G and T indicate adenine, cytosine, guanine, and thymine, respectively. Based on computer analysis (DNASIS, Hitachi Software Engineering America, Ltd., CA), this sequence was determined to be homologous to the human  $\beta_2$  adrenergic receptor mRNA.

10 5'-NH<sub>2</sub>-ATG CTG GCC GTG ACG CAC AGC A-3' (Seq ID NO: 6)

(2) Immobilization of the First Nucleotide Probe Onto Insoluble Supports

Both EDC and sulfo-NHS (Pierce, IL) were dissolved in DEPC-treated water at concentrations of 20 mM and 10mM, respectively. EDC/Sulfo-NHS solution was then prepared by mixing equal volumes of both EDC and sulfo-NHS. The first nucleotide probe was dissolved in DEPC-treated water at a concentration of 1 $\mu$ g/ $\mu$ l and then mixed with the EDC/Sulfo-NHS solution in the ration 1:25 (Vol:Vol).

50 $\mu$ l of this probe solution was added to each well of a microtiter plate (MS-3796F, Sumitomo Bakelite, JAPAN), which were known to express carboxyl groups on their surface. After incubation at room temperature overnight, the reaction solution was removed by aspiration.

(3) Preparation of mRNA

DNA plasmid pUC18 containing the human  $\beta_2$  adrenergic receptor cDNA (pTF, American Type Culture Collection) was digested with *Pst*I and *Eco*RI, and mixed with the pBC phagemid (Stratagene, CA) which had been predigested with the same restriction enzymes. The resultant plasmid was linearized by *Pst*I, extracted once with phenol-chloroform, precipitated in ethanol, and used to transcribe the human  $\beta_2$  receptor mRNA using the Riboprobe system and T7 RNA polymerase (Promega, WI). The transcribed human  $\beta_2$  receptor mRNA was treated with RNase-free DNase to digest the remaining template cDNA, then purified with one cycle of phenol extraction and ethanol precipitation. The resultant mRNA showed up as a single band on agarose (FMC bioproducts, ME) gel electrophoresis. Messenger RNA concentration was measured by a spectrophotometer at 260 and 280 nm against standard mRNA solutions prepared at concentrations of 25 pg, 250 pg, 2.5 ng, 25 ng and 250 ng, respectively.

(4) Hybridization of the First Nucleotide Probe to mRNA

In order to remove RNase from the well, the first nucleotide probe-immobilized wells were treated with 250  $\mu$ l of lysis buffer containing 0.5 M NaCl at 45°C for one hour. The buffer was removed from individual wells by aspiration, and 50  $\mu$ l of lysis buffer containing varying concentrations of standard human  $\beta_2$  receptor mRNA was added into each well. These solutions were incubated at 59°C for one hour to allow hybridization.

(5) Preparation of the Second Nucleotide Probe

The second nucleotide probe was a 15-mer oligo(dT) labeled with biotin at the 5' end.

(6) Hybridization of the Labeled Second Nucleotide Probe to the Poly(A) tail of mRNA

After hybridization of the first nucleotide probe and mRNA described in step (4), the hybridization solution was removed by aspiration and the individual wells were washed with 250  $\mu$ l of lysis buffer. Fifty  $\mu$ l of lysis buffer containing 0.5 M NaCl and 1  $\mu$ l of the biotinylated second probe solution (35 pmol/ $\mu$ l) was added into each well, and incubated at room temperature for an additional one hour.

(7) Measurement of Chemical Activities of the Labeled Second Nucleotide Probe

Following the hybridization described in step (6), the hybridization solution was removed by aspiration and wells were washed once with 250  $\mu$ l of lysis buffer. A blocking buffer consisting of (0.05%(w/v) Tween 20, 500 mM NaCl, 100 mM Tris-HCl, pH 7.5) was added into each well and incubated at room temperature for five minutes to reduced nonspecific binding. Reaction solutions were then removed by aspiration, 50  $\mu$ l of blocking solution containing 1/5000 vol. of alkaline phosphatase-conjugated streptavidin solution (Clontech, CA) was added into each well, and incubated at room temperature for 30 minutes. Each well was then washed with 250  $\mu$ l of washing buffer containing 500 mM NaCl, 100 mM Tris-HCl, pH 7.5, and 50  $\mu$ l of Lumiphos 530 solution (Boehringer-Mannheim, IN) whole plates were placed on X-ray film (XAR5, Kodak, NY) in specific holders (MYC, Kodak, NY). X-ray film was exposed to a generated chemiluminescent light at room temperature for one hour, and then developed using a film processor (QX-400, Konica, Japan). Relative intensities of individual wells on X-ray films were quantified by densitometry (model 620, Bio-Rad, CA).

(8) Results

The results of Example 1 are shown in Figure 8. The inset (A) shows results of an X-ray film exposed to chemiluminescent light during the detection of mRNAs by the invented method: (B) shows a densitographic representation of the same image as in (A):

-24-

(C) exemplifies a calibration line of the mRNA in which the X-axis represents concentrations of mRNA and the Y-axis represents the relative intensity of chemiluminescence in an individual well. Wells 1, 2, 3, 4 and 5 of Figure 8 correspond to the amounts of tested mRNA at 25 pg, 250 pg, 2.5 ng, 25 ng and 250 ng, respectively. As shown in (A), mRNAs were detected in amounts over 25 pg by the invented method. The calibration line shown in (C) shows the high linearity of signal when increasing concentrations from 25 pg to 250 ng of mRNA.

#### COMPARISON 1

##### Quantification of mRNA by Northern Blot

###### 10 (1) Agarose Gel Electrophoresis of mRNA

Ten mM of sodium phosphate buffer, pH 7.0 containing 0.1% DEPC was heated to 37°C overnight and autoclaved. A 1% agarose solution was prepared using this 10 mM DEPC-treated phosphate buffer. The solution was chilled to ca. 50°C, and iodine acetate was added to 0.2% (W/V). A gel slab was formed (6.5 cm x 10 cm) with combs and placed at room temperature for one hour to allow hardening. Five µl of DEPC-treated water containing 250ng, 25 ng, 2.5 ng, 250 pg or 25 pg of mRNA described in example 1.(3), were mixed with 7 µl of loading buffer (125 µl DMSO, 42 µl 40% glyoxal, 3 µl 1 M sodium phosphate buffer, pH 7.0), and incubated at 50°C for one hour. Samples were applied to the gel, and electrophoresed in a solution of DEPC-treated 10 mM phosphate buffer at constant 6V/cm for 15 minutes without circulation. An additional one hour of electrophoresis was performed with circulation.

Following electrophoresis, gels were immersed in 50mM sodium hydroxide containing 0.5 µg/ml ethidium bromide (Sigma, MO) for 20 minutes, and 0.1 M Tris-HCl buffer, pH 7.5 for 40 minutes. Gels were then transferred into fresh 0.1 M Tris-HCl buffer and photographed under UV light.

(2) Northern Transfer of mRNA Onto Membranes

Filter paper (3MM Chr, Whatman, NJ) and a nylon membrane (Magnagraph, Micron Separations, MA) were immersed in sterile water for 20 minutes and then placed in 10x SSPE (1.5 M NaCl, 115 mM sodium phosphate, 11.5 mM EDTA in water, pH 7.4 adjusted with 10 N NaOH). A blotting sponge (Stratagen, CA) was also immersed in 10x SSPE. The 3MM filter, nylon membrane, agarose gel, and sponge were assembled from the bottom in a Posi-blot Pressure blotter (Stratagen, CA), to cause the mRNA to be transferred onto the nylon membrane by positive pressure (835 mmHg for 45 minutes). The nylon membrane was dried, and the mRNA was permanently immobilized onto the nylon membrane by UV light cross linkage at 120 mJoules (Stratalinker 1800, Stratagen, CA).

(3) Preparation of Biotinylated cDNA Probe

DNA probes were prepared from  $\beta_2$  receptor cDNA by using a DNA labeling kit (Boehringer-Mannheim, IN) and biotin-labeled deoxyuridylic acid (Clontech, CA). In brief, a  $\beta_2$  receptor cDNA-containing plasmid, pTF was digested with *Pst*I and *Eco*RI. The resultant plasmid and cDNA were separated by agarose gel electrophoresis. A band of  $\beta_2$  receptor cDNA was cut out and removed from the gel using the Gene clean kit (Bio 101, CA). 0.1  $\mu$ g of above cDNA dissolved in 5  $\mu$ l of water was incubated at 95°C for 10 minutes, and then incubated with 2  $\mu$ l of hexanucleotide mixture, 2  $\mu$ l of dNTP and 1  $\mu$ l of Klenow enzyme (DNA labeling kit, Boehringer-Mannheim, IN) at 37°C for one hour. Following incubation, the reaction was terminated by addition of 2  $\mu$ l of 0.2 M EDTA, pH 8.0, 2.5  $\mu$ l of 4.0 M LiCl and 75  $\mu$ l of 100% ethanol, and cooled at -70°C for 30 minutes to precipitate cDNA. After centrifugation, the pellet was washed once with 70% ethanol and then dissolved in 50  $\mu$ l of water.

(4) Hybridization of Biotinylated cDNA Probe to mRNA

The mRNA containing membranes were incubated in hybridization buffer containing 0.5 M NaCl (ECL, Amersham, IL) at 45°C for one hour. Biotinylated cDNA probe described above were first heat denatured at 95°C for 10 minutes, immediately cooled on ice, and then added to the hybridization buffer. Hybridization was continued at 40°C overnight.

(5) Chemical Activity Measurements of Labeled cDNA Probe

After hybridization, membranes were washed twice in 0.5x SSPE solution containing 36% (W/V) urea, 0.4% (W/V) SDS and 0.5 M NaCl at 65°C for 15 minutes each, followed by washing twice in 2x SSPE with 0.5 M NaCl at room temperature for five minutes each. Membranes were then incubated with 100 mM Tris-HCl buffer, pH 7.0, containing 2%



-26-

blocking reagent (Boehringer-Mannheim, IN) and 150 mM NaCl at room temperature for three hours. 1/5000 vol. of alkaline phosphatase-conjugated streptavidin (Clontech, CA) was added to the above solution and incubated at room temperature 30 minutes. Membranes were washed twice with 100 mM Tris-HCl, pH 7.0, containing 150 mM NaCl at room temperature for 30 minutes each followed by one washing with alkaline phosphate buffer (100 mM NaCl, 5 mM MgCl<sub>2</sub> in Tris-HCl buffer, pH 9.5) at room temperature for 2 minutes. Membranes were then soaked in Lumi-Phos 530 (Boehringer-Mannheim, IN), and quickly placed between two sheets of transparent plastic film. Chemiluminescent membrane signals were exposed to X-ray film (XAR5, Kodak, NY) in specific holders (MYC, Kodak, NY) at room temperature for one hour.

(6) Results

The Northern blot results of Comparison 1 are shown in Figure 8(I), in which the chemiluminescent signals of mRNA were exposed to an X-ray film. Lanes 1, 2, 3, 4, and 5 of Figure 8(I) correspond to 25 pg, 250 pg, 2.5 ng, 25 ng and 250ng of applied mRNA, respectively. Because this cDNA probe is approximately 100 fold larger than the oligonucleotide probe used in Example 1, the amount of incorporated biotin is also 100 fold more than the oligonucleotide. This result indicates that the cDNA probes used in Comparison 1 are approximately 100 fold more sensitive than the oligonucleotide probe. However, as seen in Figure 8(I), a conventional Northern blot can only detect samples containing greater than 25 ng of mRNA. This is far less sensitive than the invented method as exemplified by Figure 8(I). Also, the invention as described by Figure 8(I) used less sensitive oligonucleotide probes when compared to the cDNA probes used on the Northern Blot.

EXAMPLE 2

Detection of Human  $\beta_2$  Receptor mRNA From Unknown

Samples And Evaluation of Reproducibility

(1) Preparation of the First Nucleotide Probe and Its Immobilization Onto the Insoluble Support

The first nucleotide probe was prepared and immobilized onto an insoluble support as described in Example 1.

(2) Preparation of Standard mRNA

mRNA was prepared as described in Example 1.

(3) Preparation of Cell Lysates

Human HL60 promyelocytic leukemia cells were collected by centrifugation, washed  
5 with PBS (8 g NaCl, 0.2 g KCl, 1.44 g  $\text{Na}_2\text{HPO}_4$  and 0.24 g  $\text{KH}_2\text{PO}_4$  in 1 liter of water, pH 7.4). Cells were then resuspended in lysis buffer at a concentration of  $1 \times 10^7$  cells/ml. The solution was repeatedly passed through a sterile plastic syringe fitted with a 21 gauge needle and incubated in a slow-shaking waterbath at 45°C for one hour.

(4) Hybridization and Chemiluminescent Detection

10 As described in Example 1, hybridization of the first nucleotide probe, preparation of the second nucleotide probe, hybridization of the labeled second nucleotide probe to the poly(A) tail of mRNA, and chemical activity measurement of the labeled second nucleotide probe was carried out.

(5) Preparation of Calibration Line

15 As described in Example 1, a calibration line of known  $\beta_2$  receptor mRNA concentrations was created (Figure 9). The calibration line expresses a linear co-relationship between mRNA concentrations mapped to a logarithmic scale and the relative chemiluminescent intensity read from X-ray film.

(6) Evaluation of Reproducibility

20 Concentrations of  $\beta_2$  receptor specific mRNA from HL60 cell lysates were determined three times. In Table 2, each concentration of  $\beta_2$  receptor specific mRNA was determined from the calibration line described in section (5) above. The mean and standard error (S.E.) of data were calculated. Cell numbers were counted by a phase hemacytometer (American Scientific Products, IL). As shown in Table 2, three separate assays exhibited  
25 similar mRNA values at a range of 0.96 to 1.12 fg/cell. This suggests that the invented method provides reproducible method of quantifying specific mRNA levels.

-28-

TABLE 2

Concentrations of  $\beta_2$  Receptor-Specific mRNA in HL60 Cells

No. of assay	mRNA Concentration [tg/cell]
1	1.01
2	0.96
3	1.12
mean $\pm$ S.E.	1.03 $\pm$ 0.04

G Proteins

Cell surface receptors for hormones and neurotransmitters are known to be coupled to intracellular heterotrimeric GTP-binding proteins (G proteins) composed of  $\alpha$ ,  $\beta$  and  $\gamma$  subunits. Once receptors are activated by specific ligands, receptor-coupled G proteins transduce signals to intracellular secondary effector systems, such as adenylyl cyclase, phospholipase C, and ion channels.

Recently, the molecular mechanisms of the interaction between receptors and G proteins have become less clear, because extensive molecular cloning has identified more subclasses of G protein  $\alpha$  subunits than there are known cellular functions of G proteins. Furthermore, it has been found that a single receptor is also capable of coupling to multiple G proteins. For example, histamine  $H_2$  receptors in human HL-60 cells were coupled to two different G proteins, which activated adenylyl cyclase and phospholipase C, respectively. It has also been reported that GABA-B receptors were capable of coupling to  $G_0$ ,  $G_{0^*}$  and  $G_{i_1}$  proteins, but not  $G_{i_3}$  proteins. Moreover, CHO cells have been shown to express multiple G proteins, which activate the same phospholipase C pathways but are coupled selectively to different receptors. Some receptors may alter their G-protein subclass association depending on the state of differentiation of the cell or on the particular phase of the cell cycle. One example would be the mRNA coding for G protein. G protein binds to and controls the function of various receptors in the cell membrane. It catalyzes the intracellular reactions of cAMP and calcium ion free reaction systems. These molecules are used in various cellular transmitter systems.

G proteins are believed to be involved in causing various disease states. For example, a genetic deficiency of  $G_s$  proteins is the molecular basis of hereditary

osteodystrophy. Pituitary tumors in acromegalic patients have been shown to contain mutant Gs proteins. G proteins are also involved in invasive and metastatic melanoma cells. Rat models of streptozotocin-induced experimental diabetes suggest that the levels of mRNA for various subclasses of G $\alpha$  proteins are significantly altered from normal control rats. Furthermore, cellular functions of pertussis toxin-sensitive G proteins were shown to be significantly impaired in atherosclerotic porcine coronary arteries, while G protein function in leukocytes of patients with mania was hyperfunctional. Thus, those diseases which are known to be related to a malfunction of the G protein include atherosclerosis, bipolar disorder, diabetes mellitus, melanoma, and pituitary tumors.

Currently available immunological detection methods (Western blots) and mRNA detection methods (Northern blots) are not sensitive and require a lot of cellular material, making it difficult to study the role of G proteins in disease. It is not easy to ascertain the quantity of expressed G protein. If the mRNA for the G protein was measured, it could be used to recognize the pathophysiology of patients having one of the foregoing ailments. The effects of agents designed to arrest or control these conditions could also be measured using this technique. G protein consists of  $\alpha$ ,  $\beta$  and  $\gamma$  subunits. The  $\alpha$  subunit is known to have several allelic variations. Thus, by measuring each variation of G protein mRNA, the amount of expressed protein for each variation can be estimated.

### **EXAMPLE 3**

#### **Quantification of Human G Protein $\alpha$ Subunit-Specific mRNA**

##### **(1) First Nucleotide Probe Preparation**

The first nucleotide probes used against Gs protein-specific mRNA and Gi-2 protein-specific mRNA were oligodeoxyribonucleotides. These nucleotide sequences, shown below, contain an amino group on their 5' end and were synthesized by Genosys (Texas).

The probe sequence used for quantifying Gs mRNA is shown below (Seq ID No. 1):

5'-NH<sub>2</sub>-TTC ATC CTC CCA CAG AGC CTT G-3'

The sequence of probe used for quantifying Gi-2 mRNA is shown below (Seq ID No. 2):

5'-NH<sub>2</sub>-ATG GTC AGC CCA GAG CCT CCG G-3'

(2) Immobilization of the Second Nucleotide Probe Onto Insoluble Supports

The first nucleotide probes were dissolved in DEPC-treated water at  $1\mu\text{g}/\mu\text{l}$ . Probes were then mixed with the EDC/Sulfo-NHS solution described in Example 1 Paragraph (2) in a ratio of 1:25 (Vol:Vol). A  $50\mu\text{l}$  sample was added to each well of a Sumilon microtiter plate (MS-3796F, Sumitomo BAKelite, JAPAN) and incubated overnight at room temperature. Following incubation, the reaction solution was removed with an aspirator.

(3) Preparation of Standard mRNA

A plasmid vector pGEM2 containing rat Gs protein cDNA (provided by Dr. R. R. Reed, Johns Hopkins University, U.S.A.) was linearized by digestion with *NheI* (Promega, WI). Gs protein-specific mRNA was prepared by the method described in Example 1 Paragraph (3) using T7 RNA polymerase (Promega, WI). The resultant mRNA concentrations were determined by spectrophotometry and subsequently diluted with DEPC-treated water as described in Example 1 Paragraph (3).

(4) Preparation of Cell Lysates

Human B lymphocyte-derived IM9 cells were centrifuged, washed with PBS, and resuspended in lysis buffer at  $1 \times 10^7$  cells/ml as described in Example 2. Cells were then lysed by repeated passage through a sterile plastic syringe fitted with a 21 gauge needle, and then incubated in a slow-shaking waterbath at  $45^\circ\text{C}$  for one hour.

(5) Hybridization of the First Nucleotide Probe to mRNA :

In order to remove RNase contamination in the microtiter well,  $250\mu\text{l}$  of lysis buffer was added to each well at  $45^\circ\text{C}$  for one hour as described in Example 1. The buffer was aspirated, and  $50\mu\text{l}$  of lysis buffer containing  $0.5\text{ M NaCl}$  and various concentrations of human Gs protein mRNA was added into each well. Each sample was incubated at  $51^\circ\text{C}$  for one hour to allow hybridization as described in Example 1.

(6) Hybridization of the Labeled Second Nucleotide Probe to a Poly(A) Tail of mRNA, and Measurement of Chemical Activities of the Labeled Second Nucleotide Probe

After hybridization of the first nucleotide probe, the hybridization solution was aspirated off and individual wells were washed once with 250  $\mu$ l of lysis buffer. Fifty  $\mu$ l of lysis buffer containing 0.5 M NaCl and 1  $\mu$ l of the biotinylated second nucleotide probe solution (35 pmol/ $\mu$ l) was added into each well, and incubated at room temperature for one hour to allow hybridization.

Chemical activities of the labeled second nucleotide probe was measured as described in Example 1.

(7) Preparation of Calibration Line

The calibration line of standard Gs protein-specific mRNA is shown in Figure 10. The calibration line expresses a relationship between the common logarithm of mRNA concentrations and the relative intensity of chemiluminescent signals exposed to X-ray film.

(8) Evaluation of Reproducibility

Concentrations of Gs protein and Gi-2 protein specific mRNAs from IM9 cell lysates were measured six times to test reproducibility of the present invention. The methods used for Gi-2 were similar to that described above for Gs, with the incubation temperature used for hybridization of Gi-2 mRNA being 59°C. Table 3 lists the mRNA concentrations as obtained from the calibration line, and, the mean, and the standard error. Cell numbers from each sample were counted with a phase hemacytometer (American Scientific Products, IL). As listed in Table 3, all six assays reported similar mRNA concentrations for both Gs and Gi-2 mRNA. This suggests that the current invention provides a reliable, reproducible means of measuring specific mRNA concentrations.

-32-

TABLE 3

	No. of Assay		mRNA Concentration [fg/cell]	
	Gs		Gi-2	
5				
	1	24.8	4.6	
	2	10.6	8.8	
	3	18.8	7.2	
10	4	35.5	6.8	
	5	13.6	11.5	
	6	33.0	13.7	
	mean $\pm$ S.E.		22.7 $\pm$ 3.8	8.77 $\pm$ 1.24
15				

**EXAMPLE 4**

Simultaneous Quantification of Various mRNAs  
Derived From Cell Lines by the Invented Method  
in Conjunction with Reverse PCR Method

(1) Simultaneous Detection of Gs and Gi-2 Protein-Specific mRNAs in Various Cell Lysates by the Invented Method

Using the method described in Example 3, specific mRNAs for both Gs and Gi-2 protein from various cell lysates, such as Jurkat human T-cells, IM9 human B cell, U937 human monocytes and HL60 human granulocytic cells were analyzed. Chemical activities of the labeled second probe were exposed to Polaroid film (667, Polaroid, MA) in a specific holder (model 901, Tropix, CA) at room temperature for 5 seconds, instead of exposing to X-ray films as described in Example 3.

(2) Simultaneous Detection of Gs and Gi-2 Protein-Specific mRNAs by the Conventional Reverse PCR Method

(a) cDNA Preparation and Amplification by PCR

mRNA was purified using the FastTrack kit (Invitrogen, CA) directly from various cell lysates as described in Example 3 Paragraph (4). Samples were dissolved in 10  $\mu$ l DEPC-treated water, and cDNA was synthesized using the Superscript kit (Gibco/BRL Life

Technologies, MD). Resultant cDNA was extracted once with phenol, precipitated in ethanol, and dissolved in 10 µl of DEPC-treated water.

One µl of solution containing cDNA was mixed with 1 µl each of 10 mM dATP, dCTP, dGTP and dTTP (Pharmacia LKB Biotechnology, NJ), 1 µl of 25 mM MgCl<sub>2</sub>, 1 µl each of G protein-specific PCR primers, 0.5 µl of *Taq* polymerease (Promega, WI), and 5 µl of PCR buffer (Promega) to a final volume of 50 µl. Sense and antisense G protein-specific PCR primers are exhibited below

AGCACCATTTGTGAAGCAGATGA (sense) (Seq ID No. 3)

CTCTGGCCTCCCACATCAAACA (5' to 3' antisense) (Seq ID No. 4)

These sequences are known to be highly conserved among various G proteins, with this pair of primers being known to amplify Gs, Gi-1, Gi-2, Gi-3 and Go protein specific cDNA from data gathered in preliminary experiments. Followed by pre-denaturation at 95°C for 10 minutes, PCR was carried out in a DNA thermal cycler (Perkin-Elmer Cetus, CT) with 30 cycles of annealing at 55°C for 1.5 minutes, extension at 72°C for 4 minutes, and denaturation at 95°C for 1.5 minutes, respectively.

(b) Agarose gel electrophoresis of amplified cDNA

A 1.2% agarose gel containing 0.00005% ethidium bromide, 0.001 M EDTA and Tris-acetate, pH 8.0 was prepared. Ten µl of each test solution was mixed with 2 µl of gel-loading buffer (0.25% bromophenol blue, 0.25% xylene cyanol and 15% Ficoll (Type 400, Pharmacia LKB Biotechnology, NJ; in water) and heated at 65°C for five minutes prior to loading. Electrophoresis was performed at 6 V/cm for one hour. After electrophoresis, the cDNA bands were examined under UV light illumination.

The gel was then immersed in 0.25 M HCl for 30 minutes (depurination), 0.5 M NaOH containing 1.5 M NaCl for one hour (denaturation), followed by 1 M Tris-HCl, pH 7.5 containing 1.5 M NaCl for one hour (neutralization).

(c) Transfer of Amplified cDNAs onto Membranes (Southern Blot)

Amplified cDNAs were transferred onto nylon membranes and immobilized by UV light as described in Comparison 1(2).

(d) Preparation of Biotin-Labeled cDNA Probes

Ten ng each of Gs and Gi-2 cDNA clones were amplified by PCR as described in (2)(a) above, with the modification that 20% of dTTP was replaced with biotin-labeled dUTP (Clontech). In preliminary Southern blot analysis, these biotin-labeled PCR products have been shown to specifically hybridize to Gs and Gi-2, respectively.

(e) Hybridization of Labeled cDNA Probes to Southern Membranes



As shown in Comparison Example 1 Paragraph 4, biotin-labeled PCR products were heat-denatured at 95°C for 10 minutes, then hybridized to Southern membranes, on which PCR-amplified cDNAs from cell lysates were immobilized.

(f) Chemical Activity Measurement of Labeled cDNA Probes

5 Chemical activity of labeled cDNA probes was measured as described in Comparison 1 Paragraph (5).

(3) Results

10 The results of Example 4(1) are shown in Figure 11. Polaroid film was exposed to chemiluminescent light derived from an oligo-(dT) probe by the invented method, in which Gi-2 and Gs protein mRNA from various cells were trapped onto the first oligonucleotide-immobilized plastic plate. Phosphate-buffered saline (PBS) was used as a control. As shown in Figure 11, mRNA was detected from all cell lines.

15 The results of Example 4(2) are shown in Figure 12, in which Gi-2 and Gs cDNA probes were used on a Southern Blot. PCR-amplified G protein specific DNA from four different cell lines were immobilized onto a membrane. Figure 12 is a copy of the X-ray film detailing the chemiluminescent results of this probe. As shown in Figure 12, mRNAs are detected in all cell lines. Therefore, the invented method provides a high enough reliability to be used for quantification of mRNA.

EXAMPLE 5

20 Methods for Amplifying G Proteins With PCR Primers

(1) Materials

25 The cDNAs of rat G protein  $\alpha$  subunits ( $G_{i-1}$ ,  $G_{i-2}$ ,  $G_{i-3}$ ,  $G_s$  and  $G_o$ ) were provided by Dr. R.R. Reed (Johns Hopkins Univ., MD).  $\lambda$ ZAP libraries of rat pituitary and intestine were provided by Dr. D.G. Payan (Univ. Calif. San Francisco). Kirsten murine sarcoma virus transformed rat kidney cells (KNRK), human IM9 B-lymphocytes and human Jurkat T-lymphocytes were obtained from American Type Tissue Culture Collection, Rockville, MD). Cell culture media, Superscript (Gibco/BRL, Gaithersburg, MD), reagents for PCR (Promega, Madison, WI), ECL (Amersham, Arlington Height, IL), Genius, Lumi-Phos 530 (Boehringer-Mannheim, Indianapolis, IN), FastTrack (Invitrogen San Diego, CA),  $\lambda$  gt10 library of human HL-60 cells, biotin-dUTP, alkaline phosphatase-conjugated streptavidine (Clontech, Palo Alto, CA), dNTP (Pharmacia, Piscataway, NJ) were obtained from the designated suppliers. Other chemicals were purchased from Sigma (St. Louis, MO).

30 (2) Cell culture

KRNC cells were grown in Dulbecco's modified Eagles medium containing 10% fetal calf serum, 100 U/ml penicillin and 100 µg/ml streptomycin at 37°C in 5% CO<sub>2</sub>/95% air. Cells were fed every other day and passaged at 70-90% confluency with 0.1% trypsin in CA<sup>2+</sup>-Mg<sup>2+</sup>-free saline containing 0.02% EDTA. IM9 and Jurkat cells was grown in RPMI 1640 containing 10% fetal calf serum, 100 U/ml penicillin and 100 µg/ml streptomycin at 37°C in 5% CO<sub>2</sub>/95%. Cell viability was more than 90% as assessed by the exclusion of trypan blue.

(3) **Primer design**

Rat clones of Gα proteins (G<sub>i-1</sub> (RATBPGTPB), G<sub>i-2</sub> (RATBPGTPA), G<sub>i-3</sub> (RATBPGTP), G<sub>s</sub> (RATBPGTPD), and G<sub>0</sub> (RATBPGTPC)) were retrieved from GenBank release 65.0 (HIBIIO, Hitachi America, Brisbane, CA). The nucleotide sequence similarity among these clones were then analyzed by the multiple alignment program (DNASIS, Hitachi). We have initially identified 7 highly conserved areas among them. These conserved nucleotide sequences were then analyzed against all mammalian sequences in GenBank in order to identify other similar sequences. The designed oligonucleotides (Seq. ID No. \*) were synthesized by Genosys Biotechnologies (Woodlands, TX), and suspended in water at 100 pg/ml.

(4) **PCR**

One µl of the template DNA was mixed with 1 mM each of dATP, dGTP, dCTP and dTTP, 1 µl of each PCR primers, 1 µl of 25 mM MgCl<sub>2</sub>, 5 µl PCR buffer, and 0.5 µl of Taq polymerase (18). PCR was then carried out in a DNA thermal cycler (model 480, Perkin-Elmer Cetus, Norwalk, CT) with 30 cycles of annealing temperature at ranging from 37°C to 65°C for 1.5 min, 72°C extension for 4 min followed by 95°C denaturization for 1.5 min. In separate experiments, 35% of dTTP was replaced with biotin-dUTP in order to prepare biotin-labeled probes.

(5) **Southern blot**

PCR products were separated by electrophoresis in 1.2% agarose, and stained with ethidium bromide (19). Gels were then depurinated in 0.25 N HCl for 30 minutes and denatured in 0.5 N NaOH containing 1.5 M NaCl for 30 minutes. The gels were then neutralized with 1.0 M Tris, pH 7.6 containing 1.5 M NaCl for 30 minutes. Gels were then placed onto nylon membranes (MagnaGraph, MSI, Westboro, MA) prewetted in 10X SSPE for 10 min, and DNA was transferred onto membranes by positive pressure at 75 mmHg for 60 minutes (Posiblot, Stratagene, La Jolla, CA). The DNA from the gel was then cross-linked to the membranes with ultraviolet light at 120 mJoules (Stratalinker, Stratagene), and

the membranes were incubated with hybridization buffer (ECL) containing 5% blocking reagent (ECL) and 0.5 M NaCl at 40°C for more than 1 hour. Heat denatured biotin-labeled PCR probes were then added, and hybridization was continued overnight. The membranes were washed four times for 15 minutes each time with primary wash buffer (0.5x SSPE, 36 w/v% urea, 0.4w/v% SDS) at 45-65°C, then washed twice for 5 minutes with secondary wash buffer (2x SSPE) at room temperature, and were incubated with the blocking buffer (Genius) for at least 3 hours at room temperature. Alkaline phosphatase-conjugated streptavidine (1:5,000 dilution) was then added, and incubation was continued for an additional 30-60 minutes at room temperature. The membranes were washed four times for 15 minutes with buffer A (100 mM Tris, pH 7.5, 150 mM NaCl) at room temperature, were washed for 2 minutes once with buffer C (100 mM Tris, pH 9.5, 100 mM NaCl, 50 mM MgCl<sub>2</sub>), and soaked in Lumi-Phos 530 for approximately 1-2 minutes. The membranes were then wrapped with transparency films, and chemiluminescent signals were allowed to expose X-ray films (XAR-5, Kodak, Rochester, NY) for between 10 minutes and 1 hour.

(6) mRNA preparation and cDNA synthesis

The cells were washed with phosphate buffered saline three times, homogenized in lysis buffer (FastTrack), and then incubated at 45°C for 1 hour to eliminate any RNase activity. NaCl concentrations were adjusted at 0.5 M, and an oligo (dT) cellulose tablet was added to lysis buffer. Incubation was then continued at room temperature for an additional 40 minutes. After oligo (dT) cellulose was washed with binding buffer (FastTrack) four times, bound mRNA was eluted with DEPC-treated water. Concentrations of mRNA were determined in a spectrophotometer (Hitachi, U-2000, Irvine, CA) at OD<sub>260</sub>. The first strand cDNA was synthesized from a template mRNA in the presence of 50 mM Tris, pH 8.3, 75 mM KCl, 3 mM MgCl<sub>2</sub>, 10 mM DTT, 0.5 mM each of dATP, dCTP, dGTP, and dTTP, poly (dT) as a primer, and reverse transcriptase (Superscript) at 37°C for 1 hour. Second strand cDNA was then synthesized in the same tube, containing 25 mM Tris, pH 7.5, 100 mM KCl, 5 mM MgCl<sub>2</sub>, 10 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 0.15 mM β-NAD<sup>+</sup>, 250 μM each of dATP, dGTP, dCTP and dTTP, 1.2 mM DTT, 65 U/ml DNA ligase, 250 U/ml DNA polymerase, and 13 U/ml RNase H (Superscript) for 2 hours at 16°C. Synthesized cDNAs were then extracted once with an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1), precipitated with ethanol, and resuspended in H<sub>2</sub>O.

(7) Graphic presentation

Data on Polaroid films and X-ray films was scanned by Stratascan (Stratagene) with optimization of signal-to-noise ratio, then edited with desk top publishing software (PageMaker, Aldus, Seattle, WA).

### **EXAMPLE 6**

#### **Characterization of Cellular Components of G Protein $\alpha$ Subunits Using PCR**

##### **(1) Introduction**

Using the methods of Example 5, we have now identified and characterized two highly conserved oligonucleotide sequences among five different  $\alpha$  subunits of G proteins, SEQ ID NO:7 and SEQ ID NO:8, which can be used as PCR primers. We have found that the sequences of all the subclasses of G proteins can be amplified under the same PCR conditions using these two oligonucleotides as PCR primers, including both G protein sequences obtained from a mixture of rat G $\alpha$  protein clones and cDNAs derived from various human tissues.

Although G proteins have been analyzed extensively from a biochemical and immunological point of view using various antibodies, antibody production without any cross-reactivity among various subclasses or with high species specificity has been quite difficult to obtain. Therefore, recent experiments have focused on Northern blot analyses to identify G protein-specific mRNA from various tissues or cells in different species. However, Northern blots require experienced handling and protection from RNase contamination, in addition to a large amount of starting cellular materials.

In contrast to these conventional methods, PCR technology is more convenient and practically useful, because it requires less material than a Northern blot analysis and has great sensitivity. However, it is difficult for PCR to quantify the amount of DNA or mRNA in starting materials. Interestingly, the final PCR products obtained using the novel G-protein PCR primers of the present invention reflect the relative composition of each of the subclasses of G $\alpha$  proteins present in the starting materials. This is probably because the five different G $\alpha$  proteins cDNAs are amplified at a similar rate with a single set of PCR primers under the same PCR conditions. If known mixtures of each of the subclasses of G $\alpha$  protein clones are assayed together with unknown test samples, as shown in Figure 13, the relative composition of G $\alpha$  proteins can be determined fairly precisely. Therefore, the present method is ideal for the characterization of G $\alpha$  proteins in various tissues and cells.

Performing PCR with the primers of the present invention is also useful in clinical and diagnostic assays in the detection of disease. Since G protein abnormalities have been

associated with hereditary diseases, cancer, forms of diabetes, and other diseases, the present PCR primers for detecting and quantifying G proteins can be used to detect these diseases and assess their severity.

5 In the present study, we have identified  $G_{i-1}$ ,  $G_{i-2}$ ,  $G_{i-3}$ ,  $G_5$  and  $G_0$  using the PCR primers of the present invention. Although recent cloning has identified more subclasses of G proteins, all of these newly identified G protein cDNAs showed high degree of homology to other known G proteins. Therefore, it is expected that the primers of the present invention will amplify these subclasses as well, and that the present PCR the present PCR technique can also be applied to these new G proteins. Moreover, this PCR method  
10 can be utilized to clone unique G protein genes as well.

(2) Primers

We have designed two 22-mer oligonucleotides ( $G_2$  and  $G_4$ --SEQ ID NOS:700 and 701) as PCR primers. As shown in Table 4 below, these oligonucleotides contain sequences which are highly conserved among five different  $G\alpha$  protein cDNAs, having only 0-4 base  
15 mismatches per sequence. No mismatch was found in the 4 bases at the 3' end of the  $G_2$ -sense and  $G_4$ -antisense sequences. Furthermore,  $G_2$  and  $G_4$  have no self-complementary sequences more than 3 base pairs in a row (data not shown). In order to analyze whether  $G_2$  and  $G_4$  are common to all the  $G\alpha$  proteins, but not to other unrelated sequences, a homology search (DNASIS) of  $G_2$  and  $G_4$  sequences was carried out against all mammalian  
20 sequences in GenBank. As a result,  $G_2$  and  $G_4$  were found to be common to all the types of  $G\alpha$  proteins and rhodopsins of various species, but less homologous to other unrelated sequences (data not shown).

TABLE 4

Two consensus oligonucleotides (G2 and G4)  
among five different cDNAs of G protein  $\alpha$  subunits

	Consensus sequence (# of mismatch) [SEQ ID NO:]		
	G2 AGCACCATTTGTGAAGCAGATGA [700]	Length (bp)	G4 TGTTTGATGTGGGAGGCCAGAG [701]
Gi-1	AGCACATTGTGAAGCAGATGA (1) [702]	476	TGTTGACGTGGGAGGCCAGAG (1) [805]
Gi-2	AGCACCATcGTcAAGCAGATGA (2) [703]	479	TGTTTGATGTGGGtGGtCAGcG (3) [806]
Gi-3	AGtACtATTGTGAaCAGATGA (3) [704]	476	TGTTTGATGTaGGtGGCCAaAG (3) [707]
Gs	AGCACCATTTGTGAAGCAGATGA (0) [705]	524	TGTTcGATGTGGcGGCCAGcG (3) [708]
G0	AGCACCATTTGTGAAGCAGATGA (0) [706]	479	TGTTTGAcGTtGGgGGCCAGcG (4) [709]

TABLE 5

Nucleotide sequence similarity of PCR products  
among 5 different G $\alpha$  protein clones.

	Gi-1	Gi-2	Gi-3	Gs	G0
Gi-1	-	73.4	76.8	50.2	64.0
Gi-2	73.4	-	67.8	52.3	65.9
Gi-3	76.8	67.8	-	47.6	62.6
Gs	50.2	52.3	47.6	-	55.6
G0	64.0	65.9	62.6	55.6	-

### (3) PCR of rat cloned G $\alpha$ protein cDNAs

The cloned rat G<sub>i-1</sub>, G<sub>i-2</sub>, G<sub>i-3</sub>, G<sub>s</sub> and G<sub>0</sub> cDNAs were diluted at 10 ng/ $\mu$ l, and 1  $\mu$ l of DNA was mixed with 1 mM each of dATP, dGTP, dCTP and dTTP, 1  $\mu$ l of each PCR primers, 1  $\mu$ l of 25 mM MgCl<sub>2</sub>, 5  $\mu$ l PCR buffer, and 0.5  $\mu$ l of Taq polymerase as described in the Materials and Methods. PCR was then carried out with 30 cycles of annealing temperature at 45°C for 1.5 min, 72° extension for 4 min followed by 95°C denaturation for 1.5 min, 72°C extension for 4 min followed by 95°C denaturation for 1.5 min. PCR products were then separated in 1.2% agarose gel, and stained with ethidium bromide (upper panel). Marker indicates a 0.6 Kb fragment of HindIII digested  $\lambda$  DNA. DNAs were then transferred onto nylon membranes, hybridized with subclass specific biotin-labeled probes as described in the Materials and Methods. The membranes were washed with

primary wash buffer (0.5x SSPE, 36 w/v%)  
 blocking buffer for at least 3 hours at room  
 conjugated streptavidine (1:5,000 dilution) was at 65°C, incubated with the  
 an additional 30 min at room temperature. After an alkaline phosphatase-  
 soaked in Lumi-Phos 530 for approximately 1-2 min, the reaction was continued for  
 exposed to X-ray film for 10 min to 1 hr (lower 5 panels). membranes were

PCR was first carried out at different annealing temperatures from 37°C to  
 65°C using the  $\lambda$  gt10 library of human HL-60 cells. As a result, PCR  
 only at 45°C and 55°C with a size of approximately 500 bp (data not shown).  
 similar to theoretical values (576 to 524 bp) (see Table 4 above). Therefore,  $G_1$  was

As shown in Figure 14, cloned rat  $G\alpha$  protein cDNAs ( $G_{i-1}$ ,  $G_{i-2}$ ,  $G_{i-3}$ ,  $G_5$ ,  $G_0$ ) were  
 successfully amplified using the same set of PCR primers ( $G_2$  and  $G_4$ ) with a size of  
 approximately 500 bp in 1.2% agarose gels stained with ethidium bromide. According to  
 the computer analysis (DNASIS), the nucleotide sequences of the amplified PCR products  
 were less homologous among five  $G\alpha$  proteins with the percentage of similarity ranging from  
 76.8% to 47.6%. This indicates that after PCR amplification, each of the components of  $G\alpha$   
 proteins can be identified by Southern blot analysis, even though the sizes of the PCR  
 products generated are very similar among the five  $G\alpha$  proteins. Therefore, another PCR  
 was carried out in which 35% of the dTTP was replaced with biotin-conjugated dUTP in  
 order to prepare subclass-specific, biotin-labeled probes. Southern membranes were then  
 probed with these biotin-PCR products. As shown in Figure 14, these biotin-PCR probes  
 were highly specific to each  $G\alpha$  protein subclass with washing temperature at 65°C. At low  
 stringent washing, these probes cross-hybridized with other subclasses of  $G\alpha$  proteins (data  
 not shown).

#### (4) PCR of a mixture of rat cloned $G\alpha$ protein cDNAs

$G_{i-1}$ ,  $G_{i-2}$ ,  $G_{i-3}$ ,  $G_5$  and  $G_0$  cDNA were diluted at 10 ng/ $\mu$ l (+++), 100 pg/ $\mu$ l (++)  
 and 1 pg/ $\mu$ l (+), and various combinations of these diluted cDNA were used as a PCR  
 templates (upper panel). Then, PCR and Southern blot were carried out as described in  
 part (3) of this Example.

By using the  $G_2$  and  $G_4$  sequences, all the subclasses of  $G\alpha$  protein cDNA were  
 amplified with PCR when an equal amount of  $G_{i-1}$ ,  $G_{i-2}$ ,  $G_{i-3}$  and  $G_0$  were present in test  
 samples Figure 13, lane 4, 5, 10). However, if all the concentrations of  $G\alpha$  protein cDNA  
 are abundant,  $G_0$  is less amplified (Figure 13, lane 10), probably because the number of

mismatches between  $G_0$  and  $G_4$  is higher than others between  $G_4$  and the  $G_2$  sequences. If 1 or 2 of the 5  $G\alpha$  protein cDNAs were present in smaller quantities than the others, the amounts of amplified cDNA were relatively correlated with the starting concentrations of cDNAs (Figure 13, lane 1, 2, 3, 8, 9). Furthermore, if 1 of 5 of the  $G\alpha$  proteins' cDNA is more abundant than that of the others, this G protein gene was amplified more than others (Figure 13, lane 6, 7).

(5) PCR of rat cDNAs from various tissues

Rat  $\lambda$ ZAP cDNA libraries of pituitary (P) and intestine (I), and cDNA from the mRNA of rat kidney cells (KNRK) were used as PCR templates. Then, PCR and Southern blot were carried out as described in part (3) of this Example.

Using this PCR method,  $G\alpha$  protein genes were amplified not only from cloned cDNAs, but also from various rat cDNAs (Figure 15). In  $\lambda$ ZAP cDNA libraries from rat pituitary glands and cDNA from rat kidney KNRK cells,  $G_0$  was more abundant than  $G_s$ ,  $G_{i-2}$  and  $G_{i-3}$ , and  $G_{i-1}$  was undetectable (Figure 15, lane 1, 2).  $\lambda$ ZAP cDNA library of rat intestine contained more  $G_{i-2}$ ,  $G_{i-3}$ , and  $G_s$  and  $G_0$  (Figure 15, lane 3).

(6) PCR of human cDNAs

cDNAs from the mRNA of human IM9 B-lymphocytes and Jurkat T-lymphocytes were used as PCR templates. Then, PCR was carried out with annealing temperature at 45°C (1,3) or 55°C (2, 4). Southern blot were then performed as described in part (3) of this Example.

According to sequence analyses, the PCR products of rat  $G_{i-1}$ ,  $G_{i-2}$ ,  $G_{i-3}$ ,  $G_s$  and  $G_0$  sequence amplification exhibited a high degree of homology to human G protein cDNAs, see Table 6 below). Furthermore, as shown in Figure 16, PCR with a pair of  $G_2$  and  $G_4$  primers could amplify 500 bp DNA from cDNAs of human IM9 and Jurkat cells. Unlike rat cDNAs (Figure 15), both IM9 and Jurkat cells contained all the subclasses of  $G\alpha$  proteins (Figure 16). However,  $G_{i-3}$  is relatively more abundant in IM9 cells, while  $G_s$  and  $G_0$  were more in Jurkat cells than IM9 cells (Figure 16).



TABLE 6

Nucleotide sequence similarity of PCR products between  
rat and human G proteins

5		Length (bp)		No. of mismatch	% Similarity
		Rat	Human		
	Gi-1	476	476	62	87.0%
	Gi-2	479	479	44	90.8%
10	Gi-3	476	476	40	91.6%
	Gs	524	482	52	89.2%
	Go	479	479	39	91.9%

It will be evident to one having ordinary skill in the art that a variety of sequences could serve as sense or antisense primers for PCR methods or as probes for the detection of mRNA as described herein. A method for identification of such sequences that are either common to a variety of G proteins or specific to a particular species is provided hereinbelow. In the preferred embodiment of this method of identification, a computer program is used to identify the sequences. Through use of such a program, we have identified a large number of both common and specific primers and probes. Provided as Tables 7 through 13 are various sense sequences identified through the use of such a program that are useful as G protein probes and primers.

All of the sequences listed in these tables are useful within the context of the PCR methods of the present invention. The complementary antisense sequences are also useful in certain aspects of the invention. As will be known having ordinary skill in the art, for common probes that are similar, but not identical to target sequences, stringency conditions can be varied (e.g. by changes in temperature and salinity) so that such probes will hybridize or fail to hybridize with a particular target sequence. Thus, also included within the present invention are sequences that are capable of hybridizing with the same sequences as either the sense sequences listed or their anti-sense counterparts.

Additional probes for G protein also include the following:

Common G protein probes

5'-CTCTGGCCTCCCACATCAAACA-3' (SEQ ID NO: 139)

5'-TCATCTGCTTCACAATGGTGCT-3' (SEQ ID NO: 140)

Specific probes (Human & Rat common)

-43-

Gi-1	5'-GTTTTCACTCTAGTTCTGAGAACATC-3' (SEQ ID NO: 141)
Gi-2	5'-CAAAGTCGATCTGCAGGTTGC-3' (SEQ ID NO: 142)
	5'-ATGGTCAGCCCAGAGCCTCCGG-3' (SEQ ID NO: 143)
Gi-3	5'-GTCCTCACTCTCGTCCGAAGA-3' (SEQ ID NO: 144)
5	Gs 5'-GCCTTGGCATGCTCATAGAATT-3' (SEQ ID NO: 145)
	5'-TTCATCCTCCCACAGAGCCTTG-3' (SEQ ID NO: 146)
	Go 5'-CGCATCATGGCAGAAAGCAG-3' (SEQ ID NO: 147)

-44-

Table 7. G protein common primer (G2-S)

				G2-S
				AGCACCATTGTGAAGCAGATGA
5				
	Human	SEQ ID NO:		
	Gi-1	HUMGNBPAI	148	::T::A:::::::::::::
	Gi-2	HUMGIAA	149	::::::::::C::C:::::::::
10	Gi-3	HUMGIAB	150	::::::::::::::A:::::::::
	Gs	HUMGNPAS	151	:::::::::::::::::::::
	Go	HUMGOAQ01	152	:::::::::::::::::::::
	Rat			
15	Gi-1	RATBPGTPB	153	:::::A:::::::::::::
	Gi-2	RATBPGTPA	154	::::::::::C::C:::::::::
	Gi-3	RATBPGTP	155	::T::T::::::::::::A:::::::::
	Gs	RATBPGTPD	156	:::::::::::::::::::::
	Go	RATBPGTPC	157	:::::::::::::::::::::
20	Gx	RATGXA	158	::::::::::C::C:::::::::
	Highest matched sequences in GenBank			
	HUMADECYC	adenyl cyclase	159	:::::::::::::::::::::
	RATACOAlacyl-coA	oxidase	160	GC::::::::G::::::::A:::::::::
25	HUMTGASE	transglutaminase	161	CT::::::::::::::::::CC:AC

-45-

Table 8. G protein common primer (G4-AS)

				G4-AS
SEQUENCE ID NO:				TGTTTGATGTGGGAGGCCAGAG
5	Human			
	Gi-1	HUMGNBPA	162	::::::::::::::T:::::
	Gi-2	HUMGIAA	163	::::::::::::::T::T::C:
	Gi-3	HUMGIAB	164	::::::::::A::T::::A::
10	Gs	HUMGNPAS	165	::::::C::::T::::C:
	Go	HUMGOAQ01	166	::::::C::C::::::::::C:
	Rat			
	Gi-1	RATBPGTPB	167	::::::C:::::::::::::
15	Gi-2	RATBPGTPA	168	::::::::::::::T::T::C:
	Gi-3	RATBPGTP	169	::::::::::A::T::::A::
	Gs	RATBPGTPD	170	:::C::::::::::C::::C:
	Go	RATBPGTPC	171	::::::C::T::G::::C:
	Gx	RATGXA	172	::G:G::::::::::::G::::
20	Highest matched sequences in GenBank			
	HUMLDLRRL	LDL-receptor	173	CTGA::::::::::::TACT
	MUSHEPGFA	hepatocyte growth	174	GAG:G::::::::::T::A::::
25	HUMKEREP	epidermal keratin	175	AG:::::AT:::GG:::::

Table 9. Gi-1 protein specific, human-rat common primer (Gi-1)

5				Gi-1
	SEQUENCE ID NO:			GATGTTCTCAGAACTAGAGTGAAAAC
10	Human			
	Gi-1	HUMGNBPAI	176	.....
	Gi-2	HUMGIAA	177	TT:.....CT:CCCCTGTCCCCT
	Gi-3	HUMGIAB	178	.....TC:G::G:.....G::
	Gs	HUMGNPAS	179	AC::G:C::T::TCCTG::C::G
	Go	HUMGOAQ01	180	::CA:C::C::C::G::C::
15	Rat			
	Gi-1	RATBPGTPB	181	.....
	Gi-2	RATBPGTPA	182	.....G::GC:G::CC:T::G::
	Gi-3	RATBPGTP	183	.....TC:G::G:.....G::
	Gs	RATBPGTPD	184	A:GCACAATTA:TTA:.....CG
	Go	RATBPGTPC	185	::CA:C::C::C::G::C::
20	Gx	RATGXA	186	TCA:::GAG::C::A:CC:::CA

Table 9A. Human-rodent common Gi-1 specific probes.

5				
	Name in GenBank	Gi1-735 (5'-3')	Gi1-1131 (5'-3')	SEQ ID NOS: (respectively)
10	Rat			
	Gs RATBPGTPD	A-GCACAATTA-TTA-----CG	AACAGAAA-A----AA-A---G---T	224, 225
	Gi-1 RATBPGTPB	-----	-----	226, 227
	Gi-2 RATBPGTPA	-----G--GC-G--CC-T-----G--	A-CA-----G-----G-----C-	228, 229
15	Gi-3 RATBPGTP	-----TC-G--G-----G--	--C-----T--G--CCG---	230, 231
	Go RATBPGTPC	--CA-C---C---C--G--C-----	ACA-----AG-AAA--CCGCTC	232, 233
	Human			
	Gi-1 HUMGNBPA	-----	-----	234, 235
	GenBank**			
20	HUMCPH192	C---C---G--C---G---A-----G		236
	MUSGBPA		A-C---CC--T--C-----	237

\* : Size of PCR products using Gi1-735 and Gi1-1131 as primers.

\*\* : The highest matched sequences in GenBank (release 68.0)  
next to the G protein DNAs.

25

-48-

Table 10. Gi-2 protein specific, human-rat common primer (Gi-2)

				Gi-2
5	SEQUENCE ID NO:			GCAACCTGCAGATCGACTTTG
<hr/>				
	Human			
	Gi-1	HUMGNBPAI	187	:G:GGT::A::::A:::::
	Gi-2	HUMGIAA	188	: :::::

Table 10A. Human-rodent common Gi-2 specific probes.

	Name in GenBank	Gi2-742 (5'-3') AAGATGTTTGATGTGGGTGGTC	Gi2-1102 (5'-3') AAGGAGATCTACACGCACCTTCA	SEQ ID NOS: (respectively)
5				
10	Rat			
	Gs RATBPGTPD	G-C-A-A-----ATCAAG-	---C--G-----CG-GC-ACGC	238, 239
	Gi-1 RATBPGTPB	--A-----C-----A--C-	-----A--T-----C-----	240, 241
	Gi-2 RATBPGTPA	-----	-----	242, 243
	Gi-3 RATBPGTP	--A-----A-----C-	-----G-----T-----T-	244, 245
	Go RATBPGTPC	-G-C-----C--T--G--C-	CTA-----G---C-CTCA-C	246, 247
15	Mouse			
	Gi-2 MUSGI	-----	-----	248, 249
	Human			
	Gi-2 HUMGIR	-----	-----	250, 251
20	GenBank**			
	HUMCHPF	-----C-----TGG		252
	MUSHSP84A		-----A-----A-A-G---T	253

- 25
- \* : Size of PCR products using Gi2-742 and Gi2-1102 as primers.
- \*\* : The highest matched sequences in GenBank (release 68.0) next to the G protein DNAs.



-50-

Table 11. Gi-3 protein specific, human-rat common primer (Gi-3)

5				Gi-3
	SEQUENCE ID NO:			TCTTCGGACGAGAGTGAAGAC
10	Human			
	Gi-1	HUMGNBPAI	198	:::CA:A::T:::::A::
	Gi-2	HUMGIAA	199	G::A:::::CC:C::A::::
	Gi-3	HUMGIAB	200	:::~::~~::~~::~~::~~::
	Gs	HUMGNPAS	201	GGGAAATCGA:::T:::G::
15	Go	HUMGOAQ01	202	:::C::::GA:::CGTG:A::
	Rat			
	Gi-1	RATBPGTPB	203	:::CA:A::T:::::A::
	Gi-2	RATBPGTPA	204	G::G:::::CC:T:::::
	Gi-3	RATBPGTP	205	:::~::~~::~~::~~::~~::
	Gs	RATBPGTPD	206	TTCCT::::A:::T:::T:TG
	Go	RATBPGTPC	207	CGCAT:::G::C:C::::CCA
20	Gx	RATGXA	208	AAC::::G:A::::CACCAT:

-51-

Table 11A. Human-rodent common Gi-3 specific probes.

5				
	Name in GenBank	Gi3-407 (5'-3')	Gi3-730 (5'-3')	SEQ ID NOS: (respectively)
10	Rat			
	Gs RATBPGTPD	CC-CAAGT-GATC-----TC	A-T-AT-----T--C--C----	254, 255
15	Gi-1 RATBPGTPB	A----GA-----GGC-CTA-T	--A--C-----T--C--C----	256, 257
	Gi-2 RATBPGTPA	CCTACAC---A--AT-----C	-----T--C--G--C--C----	258, 259
20	Go RATBPGTIP	-----	-----	260, 261
	Go RATBPGTPC	CG-----T--AGTC-TTAC-	TC-----AT----TCAACGAC	262, 263
	Human			
	Gi-3 HUMGTPBP	-----	-----	264, 265
	GenBank**			
	HUMAPOB	G-----C--A-----T-----C		266
	HUMINSR02		C-----G---AC---T---	267

\* : Size of PCR products using Gi3-407 and Gi3-730 as primers.

\*\* : The highest matched sequences in GenBank (release 68.0)  
next to the G protein DNAs.

-52-

Table 12. Gs Protein specific, human-rat common primer (Gs)

5				Gs
	SEQUENCE ID NO:		AATTCTATGAGCATGCCAAGGC	
	<hr/>			
	Human			
	Gi-1	HUMGNBPAI	209	GGGCGG::::T::C::::CT
10	Gi-2	HUMGIAA	210	:::ATG::::GCA::::GCTA
	Gi-3	HUMGIAB	211	C:G::::ACTA:::T::::CTC
	Gs	HUMGNPAS	212	::::::::::

15

-53-

Table 12A. Human-rodent common Gs specific probes.

5	Name		Gs-246 (5'-3')	Gs-824 (5'-3')	SEQ ID NOS: (respectively)
	in GenBank		GCCAACAAAAAGATCGAGAAGC	GGACAAAGTCAACTTCCACATG	
10	Rat				
	Gs	RATBPGTPD	-----	-----	268, 269
	Gi-1	RATBPGTPB	CG--G---G-T-----CCGCA	-----G-CGGC-G-GG-GCGC	270, 271
	Gi-2	RATBPGTPA	CAA-GG-----CA-A-----	A-----C-C--C-G-T---T	272, 273
	Gi-3	RATBPGTP	TGAGGA-----T-AA-----T	-TCAGG---A-TA-----GC	274, 275
15	Go	RATBPGTPC	TTTGG-G-G-----TA-----T	CCGAGG--CGGCGA-G-----	276, 277
	Mouse				
	Gs-1	MUSGS	-----	-----	278, 279
	Gs-2	MUSGTPAMU	-----	-----	280, 281
	Human				
20	Gs-1	HUMGSA1R	-----	-----	282, 283
	Gs-2	HUMGSA2R	-----	-----	284, 285
	GenBank**				
		HUMMBP2	A----A--C-----A-----		286
		RATAFPM		A-----T---T-----A-	287

\* : Size of PCR products using Gs-246 and Gs-824 as primers.

\*\* : The highest matched sequences in GenBank (release 68.0) next to the Gs protein DNAs.

-54-

Table 13. Go Protein specific, human-rat common primer (Go)

				Go
5		SEQUENCE ID NO:		CTGCTTTCTGCCATGATGCG
	<hr/>			
	Human			
	Gi-1	HUMGNBPAI	213	GAC:::::TAA:::TGACA
	Gi-2	HUMGIAA	214	:::TGA:::C:A:G:C:::
10	Gi-3	HUMGIAB	215	:::A::::AGA:CTTCACA
	Gs	HUMGNPAS	216	TCAACGA:::::G:::CATC
	Go	HUMGOAQ01	217	:::~::~~::~~::~~::~~::~~::
	<hr/>			
	Rat			
15	Gi-1	RATBPGTPB	218	:::A::::CCA:G:T:::TT
	Gi-2	RATBPGTPA	219	T::::GA:CC:::GCG:::
	Gi-3	RATBPGTP	220	:::~::~G:::T:::TG:CAC
	Gs	RATBPGTPD	221	:::~::~G:::AG:A:G:C:T
	Go	RATBPGTPC	222	:::~::~~::~~::~~::~~::~~::
20	Gx	RATGXA	223	:A:::C:T:::C:::CTG:
	<hr/>			

-55-

Table 13A Human-rodent common Gc specific probes.

	Name in GenBank	Go-1224 (5'-3')	Go-1397 (5'-3')	SEQ ID NOS (respectively)	
5		AGGACATCCTCCGAACCAG	TATGACCAGGTGCTCCACG		
	<hr/>				
	Rat				
	Gs	RATBPGTDP	-----AAAAAC-A-CT	AG-----AC--G-TTC	288, 289
	Gi-1	RATBPGTPB	G--G-CG-----GG----	-----T---T--TGCT-	290, 291
10	Gi-2	RATBPGTPA	CCAT-----TTCCT--A	-TCTGTT-----GG-GA	292, 293
	Gi-3	RATBPGTF	G-CCA-AGA-----GGA	A-----TA-T-AATTT	294, 295
	Gc	RATBPGTFC	-----	-----	296, 297
	Mouse				
	Go-A	MUSGOASA	-----	-----	298, 299
15	Go-B	MUSGOASB	-----	-----	300, 301
	Hamster				
	Gc-1	HAMHITAO1	-----	-----	302, 303
	Gc-2	HAMHITAO2	-----	-----	304, 305
	Human				
20	Go	HUMGOA1(2)	-----	-----	306, 307
	GenBank**				
	MUSPC3AA	G-----T--TC-----			308
	RATACACA		-----A-GT		309

25

\* : Size of PCR products using Go-1224 and Go-1397 as primers.

\*\* : The highest matched sequences in GenBank (release 68.0)  
next to the G protein DNAs.

### Substance P

Another example of clinically useful measurement of mRNA would be for the mRNA of substance P. Substance P is a neurotransmitter. Because the nerves expressing substance P are involved in pain receptor pathways, the ability to measure the mRNA of substance P would be useful in the development of analgesics.

### EXAMPLE 6

#### Detection of Human Substance P Receptor mRNA From Substance P Receptor cDNA-Transferred Cells, Vector-Transfected Cells, and Evaluation of Their Reliability

##### (1) Preparation of the First Nucleotide Probe

The first nucleotide probe used for quantification of substance P receptor-specific mRNA was an oligodeoxyribonucleotides containing a 5' amino group. This probe was manufactured using a DNA synthesizer (model 380B, Applied Biosystems, CA) and is shown below:

5'-NH<sub>2</sub>-GGA CTT ATG AGA AAG CGT ACC A-3' (Seq ID NO:5).

##### (2) Immobilization of the Second Nucleotide Probe Onto Insoluble Supports

The first nucleotide probe were dissolved in DEPC-treated water at a concentration of 1 µg/µl, then mixed with the EDC/Sulfo-NHS solution described in Example 1 Paragraph (2) in a ratio of 1:25 (Vol:Vol). 50 µl of the above mixed solution was added to each well of a Sumilon microtiter plate (MS-3796F, Sumitomo Bakelite, JAPAN) and the plates were incubated overnight at room temperature. Following incubation, the reaction solution was removed with an aspirator.

##### (3) Preparation of Standard mRNA

A plasmid vector pKR2 containing human substance P receptor cDNA (provided by Dr. S. Nakanishi, Kyoto University, Japan) was linearized by SmaI (Promega, WI) digestion. Substance P receptor-specific mRNA was prepared by the method described in Example 1 Paragraph (3) using T7 RNA polymerase (Promega, WI). Concentrations of the resultant mRNA were determined by spectrophotometry and then diluted with DEPC-treated water to prepare standard mRNA solutions as described in Example 1 Paragraph (3).

##### (4) Preparation of Cell Lysates

One line of human T lymphocyte-derived Jurkat cells which had been transfected with substance P receptor cDNA was examined for transcription of substance P receptor mRNA (hereafter called JSP, provided by Dr. D. G. Payan, University of California, San Francisco, CA). Another line of Jurkat cells which had been transfected with a vector that

didn't express substance P receptor (hereafter called JVec, a kind gift of Dr. D.G. Payan, ibid) was used as a negative control. Cells were collected by centrifugation, washed with PBS, and resuspended in lysis buffer at  $1 \times 10^7$  cells/ml as described in Example 2. The cells were sheared by repeated passage through a 21 gauge needle, then incubated in a slow-shaking waterbath at 45°C for 1 hour.

(5) Hybridization of the First Nucleotide Probe to mRNA

In order to remove possible RNase contamination in the microtiter wells containing the bound first nucleotide probe, 250  $\mu$ l of lysis buffer was added and incubated at 45°C for 1 hour as described in Example 1. The buffer aspirated off, and 50  $\mu$ l of lysis buffer containing 0.5 M NaCl and various concentrations of standard human substance P receptor-specific mRNA was added into each well, and incubated at 50°C for 1 hour to allow hybridization as described in Example 1.

(6) Hybridization of the Labeled Second Nucleotide Probe to a Poly(A) Tail of mRNA, and Measurement of Chemical Activities of the Labeled Second Nucleotide Probe

After hybridization of the first nucleotide probe to bound mRNA as described in Paragraph (5) above, the hybridization solution was removed by aspiration and individual wells were washed once with 250  $\mu$ l of lysis buffer. Five  $\mu$ l of lysis buffer containing 0.5 M NaCl and 1  $\mu$ l of the biotinylated second nucleotide probe solution (35 pmol/ $\mu$ l) was added into each well and incubated at room temperature for an additional 1 hour to allow the second hybridization as described in Example 1.

Chemical activities of the labeled second nucleotide probe was also measured as described in Example 1.

(7) Preparation of Calibration Line

The calibration line of standard substance P receptor-specific mRNA is shown in Fig.

(8) Evaluation of Reproducibility

Concentrations of substance P receptor-specific mRNA in JSP and JVec cell lysates were measured twice. Table 7 shows concentrations of mRNA obtained from the calibration line as described in Paragraph (7). Cell numbers from each sample were counted by a phase hemacytometer (American Scientific Products, IL). As shown in Table 14, substance P receptor-specific mRNA concentrations were found in JSP cells that had been transfected with substance P receptor cDNA, but were not detected from JVec cells transfected with only a control vector. This shows that the invented method provides good quantification of a specific mRNA with high reliability.



TABLE 14

Concentrations of Substance P Receptor-Specific mRNA in Jurkat Cells Transfected with Substance P Receptor cDNA or Only a Control Vector.

5

10

No. of Assay	mRNA Concentration	
	JSP [pg/cell] (Transfected)	JVec [fg/cell] (Control)
1	135	<0.1
2	62.4	<0.1

**Oligonucleotide Primers and Probes for the Detection of Jun Oncogenes**

15

One aspect of the present invention relates to DNA primers and probes for the detection of jun oncogenes, such as by PCR. Expression of the jun gene can serve as a marker for cell growth.

20

It is well known that certain oncogenes, such as jun or fos are most rapidly expressed when cells are stimulated with outside stimuli to proliferate. Therefore, detection or quantification of levels of expression of jun or fos oncogenes will be a good marker for cellular mitogenic activity.

25

Jun oncogenes are one of oncogenes first reported by Maki, et al. (Proc. Natl. Acad. Sci. USA. 84:2848-2852, 1987), and current intensive molecular cloning identified at least three subtypes of jun oncogenes, such as jun-B, c-jun and jun-D. Based on nucleotide sequence comparison, three subtypes of jun oncogenes in both humans and mice are different (Ryder K. Nathans D, Proc. Natl. Acad. Sci. USA 85:8464-8467, 1988; Ryder K. et al., Proc. Natl. Acad. Sci., USA 85:1487-1491, 1988; Hattori K. et al., Proc. Natl. Acad. Sci. USA 85:9148-9152, 1988; Schuette J. et al., Cell. 59:987-997, 1989). However, because it is still unclear how these three different jun oncogenes are involved in cell growth, one has to analyze all of the three genes during cell growth.

30

Northern blot analysis is widely accepted in order to detect specific genes (Sambrook J. et al., Molecular cloning. A laboratory manual, 2nd eds. pp. 7.39-7.52), and some investigators have already adapted this method to detect jun oncogenes (Sherman, et al., Proc. Natl. Acad. Sci. USA. 87:5663-5666, 1990; Oursler MJ et al., Proc. Natl. Acad. Sci.

USA, 88:6613-6617, 1991). In this method, mRNA is first purified from tissues/cell of researchers' interest, separated on agarose gel electrophoresis. After electrophoresis, mRNA is transferred onto membranes, and hybridized with radioactive probes to identify positive band(s) on autoradiography. However, this method is not so sensitive that it is difficult to identify specific signals if materials have only small quantity of genes. Alternatively, because reverse PCR ("Molecular Cloning") is also used to detect a wide variety of genes from different tissues/cells, this technique is also applicable for the detection of jun oncogene. In this method, mRNA is first convert to cDNA by reverse transcriptase, then specific gene fragments are amplified by PCR by using one set of primers (sense and anti-sense primers). Amplified gene can be seen on agarose gel electrophoresis in the prepense of ethidium bromide. This method is much more sensitive than the conventional Northern blot analysis, but in order to detect three different subtypes of jun oncogenes, one has to prepare six primers. Also, because of different nucleotide sequences between humans and mice, one has to prepare human and mice primers separately.

One goal we accomplished with the present invention was the design of oligonucleotide sequences for both sense and anti-sense primers that could be readily used for detecting the minuscule amounts of the jun gene found *in vivo*. We also wanted these primers to be useful for detecting both human and murine jun sequences. Thus, we selected primers which fulfilled the below conditions:

- (a) A common nucleotide sequence among three subtypes of jun oncogenes (jun-B, c-jun and jun-D), with maximum 4 base mismatches among them;
- (b) A common nucleotide sequence among humans and mice, with maximum 4 base mismatches between them;
- (c) At least 5 bases from 3' end are 100% identical without any mismatches among three subtypes and between humans and mice;
- (d) A nucleotide sequence length in both sense primer and anti-sense primer from 17 to 50;
- (e) A difference of  $T_m$  between sense primer and anti-sense primer is within 2 degree C;
- (f) Lack of complementary structure with more than 4 bases long in either the sense or the anti-sense primer;
- (g) Lack of complementary structure with more than 4 bases long between the sense and the anti-sense primer;

- (h) Both the sense and the anti-sense primer exist in the area of coding sequence;
- (i) The length of DNA to be amplified is greater than 200 bases; and
- (j) The nucleotide sequence homology of amplified genes among three subtypes of jun oncogenes is not greater than 80%, and after gene amplification by the PCR, one can characterize each subtype of jun oncogenes by Southern Blot.

DNA fragments (both sense and anti-sense primers) which satisfy the conditions as shown above were investigated and the several candidate oligonucleotides were synthesized. The test results showed the sense primer 5'-CCCTGAAGGAGGAGCCGCAGAC-3' and anti-sense primer 5'-CGTGGGTCAAGACTCTGCTTGAGCTG-3' were desirable candidates.

Table 15 shows the nucleotide sequence homology between a preferred sense primer and three subtypes of jun oncogenes in both humans and mice.

TABLE 15

Sense primer 5'-CCCTGAAGGAGGAGCCGCAGAC-3' (SEQ ID NO:310)			
Mouse	jun-B	--T-T--A-----	(SEQ ID NO:311)
	c-jun	-----A-----	(SEQ ID NO:312)
Human	jun-B	--T-C-----A-----	(SEQ ID NO:313)
	c-jun	-----T-----	(SEQ ID NO:314)

--: indicates identical base to the sense primer.

In addition, Table 16 shows the nucleotide sequence homology between the sense primers and three subtypes of jun oncogenes in both humans and mice.

TABLE 16

		Antisense primer	
		3'-GTCGAGTTCGTCTTTCAGAACTGGGTGC-5' (SEQ ID NO:315)	
5			
	Mouse jun-B	-----C-----A-	(SEQ ID NO:316)
	c-jun	-----T-----	(SEQ ID NO:317)
	jun-D	----T-----G-G-C----	(SEQ ID NO:318)
10	Human jun-B	----T----C-----	(SEQ ID NO:319)
	c-jun	----T-----	(SEQ ID NO:320)

:- indicates identical base to the sense primer.

Also, Table 17 shows nucleotide sequence homology of amplified genes between each of subtypes of jun oncogenes.

TABLE 17

		Subtypes	Homology (%)
20		jun-B & c-jun	72.0
		jun-B & jun-D	73.6
		c-jun & jun-D	75.3

Table 18 shows the homology between human and mouse subtypes.

TABLE 18

		Subtypes	Homology between human & mouse (%)
30		jun-B	93.0
		c-jun	91.9

The primer sequences can be modified in a variety of ways. For example, a primary amine residue, nucleotide sequence recognized by a restriction enzyme, or RNA promoter sequences for further modification after PCR amplification, can be added to the 5' end of the primers. These modifications are represented herein by the designation "X". Thus, for example, a modified primer as shown in Table 16, can be represented as 5'-XCGTGGGTCAAGACTTCTGCTTGAGCTG-3' can also be used, in which X represents the primary amine residue, nucleotide sequence recognized by the restriction enzyme, or RNA promoter sequences for further modification after PCR amplification.

The attachment of a restriction site at the 5' end (X) is useful for cloning of amplified genes. The attachment of RNA promoter sequences at 5' end (X) is useful in

RNA transcription and RNA transcription-based amplification. The attachment of primary amine at the 5' end is useful for coupling reaction to labeling compounds or solid supports, so that the amplified gene can be easily quantified.

5 For the above-mentioned nucleotide sequence (X) which undergoes cutting by a restriction enzyme, instead of those nucleotide sequences with higher frequency such as Eco RI or Bam HI, it is often desirable to use longer recognition sequences (e.g. 8 bases) which appear in lower frequency such as the nucleotide sequence, 5'-GCGGCCGC-3', which is recognized by NotI or the nucleotide sequence, 5'-TTAATTAA-3', which is recognized by PacI. The reason for this is the relatively common nucleotide sequences recognized by EcoRI or BamHI can result in false positive reaction during PCR amplification because of  
10 priming of primers to undesirable DNA sequences.

The above-mentioned nucleotide sequence (X) containing an RNA promoter can be any of a variety of such sequences, such as the T7, SP6 or T3 RNA promoter sequences.

15 The above-mentioned primary amine residues (X) can be added at the 5' end during oligonucleotide synthesis.

The DNA fragments for the sense primer and anti-sense primer in this invention can be easily synthesized in DNA synthesizer which are available in the market. These synthesized oligonucleotides can be purified by high pressure liquid chromatography or gel electrophoresis.

20 The test material to be analyzed is usually a total RNA or purified mRNA from cells or tissues. If desired, cells or tissues can be tested in their natural state without any pretreatment; however, in the case of drug testing, this can be accomplished by reacting the drug with cells or tissues in the test tube, or the drug can be administered (intravenous injection, subcutaneous injection, intramuscular injection, oral administration, intra-abdominal injection) to a laboratory animal, and after a set time period, cells or tissues can  
25 be removed for RNA/mRNA purification.

Methods of purification of total RNA or mRNA are well known in the art. Examples of such methods are the standard protocol described in Sambrook, et al., "Molecular Cloning, A Laboratory Manual, 2nd Ed.," Cold Spring Harbor Press, Cold Spring  
30 Harbor, NY (1989), and a commercially available kit such as FastTrack from Invitrogen (San Diego, CA). In either case, in order to avoid introducing RNase, hands should preferably be protected with vinyl gloves, and the instruments used for experiments not be touched with bare hands. Also, any glass containers to be used in the experiments should be heated prior to use at approximately 250 degrees C for at least 4 hours. Furthermore, 0.1% diethyl

-63-

pyrocarbonate (hereafter abbreviated DEPC) is added to the water which is to be used, and after incubation at 37 degrees C overnight, treated water is autoclaved. Other basic procedures are undertaken as outlined in the "Molecular Cloning," pp. 7.3-7.5. Vanadyl ribonucleoside complex can also be added to samples as described above.

5       The method of synthesis of cDNA from mRNA template using reverse transcriptase is described in the "Molecular Cloning," pp. 8.11-8.13.

For PCR amplification, synthesized cDNA is mixed with sense primer, antisense primer, 4 types of deoxynucleotides (dATP, dCTP, dGTP and dTTP), Taq polymerase, inorganic salts, and other necessary materials, and PCR reaction is undertaken in the  
10       Thermal cycler (Perkin-Elmer Cetus). A typical PCR method is described in "Molecular Cloning," pp. 14.2-14.33).

In order to analyze the amplified gene, it is appropriate to use electrophoresis. After the amplified gene undergoes electrophoresis in agarose gel, the DNA is stained with ethidium bromide. The amplified DNA band will be then visible under fluorescent light as  
15       described in "Molecular Cloning," pp. E.3-E.4. After taking photographs of the DNA band, it is also possible to quantify the intensity of each bands by scanning photographs and analysis of scanned picture by commercially available system, such as Stratascan (Stratagene, La Jolla).

Furthermore, after agarose gel electrophoresis, amplified genes can be transblotted  
20       onto membranes, and subtypes specific genes can be detected by hybridizing with labeled probes followed by exposing labeled signals, such as <sup>32</sup>P or chemiluminescence to either Polaroid films or X-ray films (Southern blot).

#### EXAMPLE 7

##### Synthesis of Sense and Anti-Sense Primers and 25       Amplification of Mouse Clones of Jun Oncogenes

The sense primer SEQ ID NO:9 (s943-2) and the anti-sense oligonucleotide SEQ  
ID NO:10 (AS1132-2) were synthesized by a synthesizer 380 B type (Applied Biosystems Co.). After treatment with ammonium hydroxide at 55°C overnight, synthesized oligonucleotides were dried in a Speed-Vac (Savant Co.), and the concentration was  
30       adjusted at 1 microgram/ml with water, and was stored at -20 degrees C until use.

1 microliter of one of the three types of mouse jun clones (jun-B, c-jun, or jun-D, obtained from ATCC), 1 microliter of sense primer, 1 microliter of anti-sense primer, 5 microliters of 10X buffer for PCR (Promega), 1 microliter of 25 mM magnesium chloride, 4 microliters of 10 mM dNTP mix, and 0.5 microliters of Taq polymerase (Promega) were

missed, and water was added up to a total volume of 50 microliters. After adding of two drops of mineral oil, PCR was undertaken using the thermal cycler, model 480. After reaction mixture was heated at 95 degree C for 10 min. PCR was carried out with the following cycles 30 times: annealing at 55 degrees C for 1.5 minutes, extension at 72 degrees C for 4 minutes, and denaturing at 95 degrees C for 1.5 minutes.

After PCR, 10 microliters of the sample were mixed with 1 microliter of 10x loading buffer (0.25% bromophenol blue, 0.25% xylene cyanol FF, and 15% Ficoll, Type 400), and electrophoresis was carried out in 1.5% agarose gel containing 5 microgram/ml ethidium bromide. After electrophoresis, the amplified DNA bands were visualized by an ultraviolet light.

As a result, it was found that in each of the cases where the clones jun-B, c-jun, and jun-D were used, a single band for the amplified DNA was observed at the position of approximately 270 bp. This indicates that the above-mentioned set of primers can recognize and amplify all three types of mouse jun oncogenes, such as jun-B, c-jun, jun-D.

(1) Alternative Synthesis of PCR primers for amplification of jun gene specific mRNA

Sense (jun-s) and antisense (jun-as) oligodeoxyribonucleotides specific for jun oncogenes were synthesized in a DNA synthesizer as described above. The nucleotide sequences were as follows:

jun-s 5'-CCCTGAAGGAAGAGCCGCAGAC-3' (Seq ID NO: 11)

jun-as 5'-CGTGGGTCATGACTTTCTGCTTGAGCTG-3' (Seq ID No: 12)

(2) PCR amplification of cDNA

To 1 µl of ss-cDNA as described in paragraph (IV), 0.1 µg (1 µl) each of jun-s and jun-as oligonucleotides, 5 µl of 10X PCR buffer (Promega), 1 µl of 25 mM MgCl<sub>2</sub>, 4 µl of 10 mM dNTP mixture (Promega), 0.5 µl Taq polymerase (Promega), and 36.5 µl of DEPC-water were mixed. Two drops of mineral oil were overlaid onto the reaction mixture to prevent evaporation. The reaction mixture was first heated at 95°C for 10 minutes, then PCR was carried out in a Thermal cycler (Model 480, Perkin-Elmer Cetus) with 30 cycles of annealing at 55°C for 1.5 minutes, extension at 72°C for 4 minutes and denaturing at 95°C for 1.5 minutes as described above.

Ten µl of the resultant PCR products were then mixed with 1 µl of 10X loading buffer, and applied to a 1.5% agarose gel containing 5 µg/ml ethidium bromide. Electrophoresis was carried out in a 1X TBE buffer at 100 V/6.5 cm width for approximately 1 hour. After electrophoresis, fluorescent DNA bands were recorded onto Polaroid film. Lanes 1-8 of Figure 17 indicate duplicate experiments of a 1:1000 dilution,

full strength (equivalent to 1 ml of blood), 1:10, and 1:100 dilution respectively of a leukocyte lysate. As shown in Figure 17, the jun oncogene was amplified from the 1:10 and 1:100 dilution displaying an estimated size of 370 base pairs. This size is similar to the 372bp theoretical value.

5

### EXAMPLE 8

#### The effect of pretreatment with EGF on expression of jun oncogenes in human mononuclear leukocytes

##### (1) Pretreatment of human leukocytes with EGF

40 ml of phosphate buffered saline (PBS) is added to 20 ml of heparinized human blood and mixed. 10 ml each of this sample is overlaid onto 3 ml of IsoLymph, then centrifuged for 30 minutes at 400 x g. After washing the pellet three times with PBS, the pellet is resuspended in 3 ml of PBS. 1 ml each of this sample is then placed in three tubes (No. 1 to No. 3). 1 ml of PBS is placed in the fourth tube as control. The total of four tubes are incubated for 10 minutes at 37 degrees C, then EGF (Epidermal Growth Factor) is placed in tube No. 1 at a final concentration of 30 ng/ml. After 15 minutes, EGF is placed in the same manner in tube No. 2, and incubated for another 5 minutes. After 5 min, RNA is extracted from all 4 tubes simultaneously. Thus, the time period of the pretreatment of human leukocytes by EGF is 20 minutes for tube No. 1, 5 minutes for tube No. 2, 0 minutes for tube No. 3.

20

##### (2) Extraction of RNA from cells

The above four tubes are taken out and undergo centrifugation by microfuge for 10 seconds. Then, the supernatant is discarded and the below described lysis buffer is added to the pellet, and after thorough mixing, this is incubated for 30 minutes at 45 degrees C.

Contents of lysis buffer:

25

10 mM EDTA pH 8.0

0.5% SDS (bacteria removed by non-bacterial filter)

0.2 M NaCl

DEPC treated water

RNA inhibitor 500 unites/ml

30

Vanadyl Complex 10 mM

Proteinase K 200 microgram/ml

##### (3) Purification of mRNA

After 5 M NaCl is added to the above cell lysates to obtain a final concentration of 0.5 M, oligo (dT) cellulose (Stratagene Co.) is added, and undergoes a reaction for 30



-66-

minutes at room temperature. Then, after washing the cellulose in 10 ml binding buffer (20 mM Tris-HCl, pH 7.6, 1 mM EDTA, 0.5 M NaCl) 5 times, 0.35 ml of DEPC treated water is added, and mRNA is eluted from the solid phase. Then, 53 microliters of 2 M sodium acetate and 2.5 times the volume of ethanol are added, and after cooling in a dry ice for 20 minutes, this undergoes centrifugation at 15,000 rpm for 20 minutes. After washing the pellet in 75% ethanol one time, this is dried, and then dissolved in 10 microliters of DEPC treated water.

(4) Synthesis of cDNA

50 mM Tris-HCl (pH 8.3), 75 mM KCl, 3 mM magnesium chloride, 10 mM DTT, 0.5 mM dNTP (dATP, dCTP, dGTP, dTTP), 50 micrograms/ml oligo (dT) primer, and 10,000 units/ml reverse transcriptase are added to 10 microliters of mRNA obtained from above to a total volume of 20 microliters, and this undergoes a reaction for one hour at 37 degrees C. After the reaction, 20 microliters of a phenol:chloroform:isoamyl alcohol mixture is added, and cooled for 20 minutes in a dry ice to precipitate the cDNA. After centrifugation for 20 minutes at 10,000 rpm, the pellet is washed one time in 75% ethanol. Then, after drying, this is dissolved in 20 microliters of autoclaved water, and stored at -20 degrees C.

(5) PCR

1 microliter each of sense primer and anti-sense primer (1 mg/ml) for jun gene amplification are added with 2 microliters of cDNA. After mixing this with 50 mM KCl, 10 mM Tris-HCl (pH 8.4), 2.0 mM magnesium chloride, 100 micrograms/ml gelatin, and 0.2 mM dNTP, 2.5 units of Taq polymerase are added (final volume is 50 ml). After reaction mixture was heated at 95 degree C for 10 minutes, PCR was carried out with the following cycles 30 times: annealing at 55 degrees C for 1.5 minutes, extension at 72 degrees C for 4 minutes, and denaturing at 95 degrees C for 1.5 minutes.

(6) Agarose gel electrophoresis

After completing PCR, 10 microliters of the reacted solution is taken and undergoes electrophoresis in the same method as described in Example 7. The results showed that in the leukocytes which did not undergo treatment with EGF (i.e., sample which underwent 0 treatment time in tube No. 3) was found to have a minimal band for the amplified DNA at the position of about 270 bp size. However, it was found that the band of amplified DNA was increased after 5 min, then returned to the basal levels within 20 min.

EXAMPLE 9

Effect of pretreatment with PHA on jun gene expression in human leukocytes

-67-

In the place of EGF, PHA (at a final concentration of 10 micrograms/ml) was utilized and the pretreatment times were set at 0 minutes, 5 minutes, 15 minutes, and 30 minutes. Other procedures were identical to those of Example 8. As a result, with the pretreatment with PHA done at 15 minutes, it was found that the band for the amplified DNA at the position of approximately 270 bp was maximized, but (the time periods) afterwards, this band decreased.

As with the other mRNAs that can be analyzed in accordance with the methods of the present invention, a variety of jun sequences could serve as sense or antisense primers for PCR methods or as probes for the detection of mRNA as described herein. A method for identification of such sequences that are either common to a variety of jun mRNAs or specific to a particular species is provided hereinbelow. In the preferred embodiment of this method of identification, a computer program is used to identify the sequences. Through use of such a program, we have identified a large number of both common and specific primers and probes. Provided as Tables 19 through 24 are various sense sequences identified through the use of such a program that are useful as jun gene probes and primers.

All of the sequences listed in these tables are useful within the context of the PCR methods of the present invention. The complementary antisense sequences are also useful in certain aspects of the invention. As will be known by those having ordinary skill in the art, for common probes that are similar, but not identical to target sequences, stringency conditions can be varied (e.g. by changes in temperature and salinity) so that such probes will hybridize or fail to hybridize with a particular target sequence. Thus, also included within the present invention are sequences that are capable of hybridizing with the same sequences as either the sense sequences listed or their anti-sense counterparts. Additional probes for jun genes include the following:

Common jun gene probes

5'-CCATGTCGATGGGGGACAGCGG-3' (SEQ ID NO: 13)

5'-CTGTTTAAGCTGCGCCACCTG-3' (SEQ ID NO: 14)

5'-GTCTGCGGCTCCTCCTTCAGGG-3' (SEQ ID NO: 15)

5'-CGTGGGTCAAGACTTTCTGCTTGAGCTG-3' (SEQ ID NO: 16)

Specific probes

B type: 5'-CACTTGGTGGCCGCCAG-3' (SEQ ID NO: 17)

C type: 5'-GAGCATGTTGGCCGTGG-3' (SEQ ID NO: 18)

Human D type: 5'-GATGCGCTCCTGCGTGT-3' (SEQ ID NO: 19)

-68-

Mouse D type: 5'-GCCTGTTCTGGCTTTTGAGGG-3' (SEQ ID NO: 20)

Table 19. Jun-common sense primer (S943-2).

	Locus			Sequences (5'-3')
5	GenBank	Pos		CCGCTGTCCCCCATCGACATGG
	Human			
	B: humjunca	1189	SEQ ID NO: 21	---G-----A-----
	C: humjuna	1981	SEQ ID NO: 22	--C-----
10	D: humjundr	943	SEQ ID NO: 23	---T---G-----
	Mouse			
	B: musjunba	1079	SEQ ID NO: 24	--TG-----A-----
	C: musjunc	1344	SEQ ID NO: 25	--C-----T-----
15	C: muscjun	1646	SEQ ID NO: 26	--C-----T-----
	C: musjun	1084	SEQ ID NO: 27	--C-----T-----
	D: musjund	927	SEQ ID NO: 28	-----G-----
	D: musjunda	782	SEQ ID NO: 29	-----G-----
20	D: musjundr	793	SEQ ID NO: 30	-----G-----
	Rat			
	C: atjunap1	1082	SEQ ID NO: 31	--CT-----
	C: ratrjg9	2984	SEQ ID NO: 32	--CT-----
25	Chicken			
	C: chkjun	1470	SEQ ID NO: 33	--C-----T--T-----
	Quail			
30	C: quljun	1186	SEQ ID NO: 34	--C-----T--T-----
	Drosophila			
	C: drojun	1038	SEQ ID NO: 35	A-CG-TAAT-----T-----
35	Highest matched sequences in EMBL			
	SDNAM2G Yeast NAM2 gene (SEQ ID NO:36)			-----A-A-----GAAT
	PRK2TRFB Plasmid PK2 trfB ope (SEQ ID NO:37)			GT-----GC-T-
	DMSYT D.melanogaster synap (SEQ ID NO: 38)			-----G-A-----UC-----

-70-

Table 20. Jun-common antisense primer (AS1132-2).

5	Sequences (5'-3')		GTCCACCGCGTCGAATTTGTC
	Locus	Pos	
10	Human		
	B: humjunca	1378 (SEQ ID NO: 39)	-----G-----G-----
	C: humjuna	2170 (SEQ ID NO: 40)	-----T-----
	D: humjundr	1132 (SEQ ID NO: 41)	-----G--C---
15	Mouse		
	B: musjunba	1268 (SEQ ID NO: 42)	--T-----G--C---
	C: musjunc	1835 (SEQ ID NO: 43)	-----T-----C---
	C: muscjun	1273 (SEQ ID NO: 44)	-----T-----C---
	C: musjun	1533 (SEQ ID NO: 45)	-----T-----C---
	D: musjund	982 (SEQ ID NO: 46)	-----GC-G-----
	D: musjunda	1116 (SEQ ID NO: 47)	-----G-----
20	D: musjundr	971 (SEQ ID NO: 48)	-----G-----
25	Rat		
	C: atjunap1	1271 (SEQ ID NO: 49)	-----T-----
	C: ratrjg9	3173 (SEQ ID NO: 50)	-----T-----
30	Chicken		
	C: chkjun	1659 (SEQ ID NO: 51)	-----A--T-----C---
35	Quail		
	C: quljun	1375 (SEQ ID NO: 52)	-----A--T-----C---
40	Drosophila		
	C: drojun	1227 (SEQ ID NO: 53)	--A-----GC-C--C---
45	Highest matched sequences in EMBL		
	MSATFA Human mRNA for ATF-a (SEQ ID NO: 54)		C-----G---A-C-----
	ECDCM E. coli dom gene (SEQ ID NO: 55)		AA-----ACCA
	ECDCMA E. coli dom (SEQ ID NO: 56)		AA-----ACCA
	OCIGKCI Rabbit Ig germine (SEQ ID NO: 57)		-----G-----G-CGA
	OCIGOS Rabbit Ig k2 L chain (SEQ ID NO: 58)		-----G-----G-CGA
	OCK1B4 Rabbit Ig kappa L (SEQ ID NO: 59)		-----G-----G-CGA
	OCIGKCG Rabbit Ig kappa2 J-C (SEQ ID NO: 60)		-----G-----G-CGA
	OCIGKC02 Rabbit Ig kappa1 J-C (SEQ ID NO: 61)		-----G-----G-CGA

-71-

Table 21. Jun-B specific probe (B-1258).

GenBank		
B1258(5'-3')		
5	CTGGCGGCCCAAGTG	
<hr/>		
Human		
	B: HUMJUNCA	----- (SEQ ID NO: 62)
	C: HUMJUNA	A-C--T---T----- (SEQ ID NO: 63)
10	D: HUMJUNDR	A-----TCCAA (SEQ ID NO: 64)
Mouse		
	B: MUSJUNBA	----- (SEQ ID NO: 65)
	C: MUSJUNC	A-T--C---T----- (SEQ ID NO: 66)
	MUSCJUN	A-T--C---T----- (SEQ ID NO: 67)
15	MUSJUN	A-T--C---T----- (SEQ ID NO: 68)
	D: MUSJUND	-----C-----CCCG- (SEQ ID NO: 69)
	MUSJUNDA	-----T-----GCCA- (SEQ ID NO: 70)
	MUSJUNDR	-----T-----GCCA- (SEQ ID NO: 71)
Rat		
20	C: RATRJG9	-----T-----GCCAA (SEQ ID NO: 72)
	RATJUNAP	A-C--T---T----- (SEQ ID NO: 73)
Chicken		
	C: CHKJUN	-C-----G-----CCC (SEQ ID NO: 74)
Quail		
25	C: QULJUN	GGC-----AG----TG- (SEQ ID NO: 75)
Drosophila		
	C: DROJUN	G----T--AT----- (SEQ ID NO: 76)
<hr/>		
Homologous sequences in GenBank		
30	J04695 Figure 2. Nucleo. G-----A-----G	(SEQ ID NO: 77)
	M27884 Figure 2. Nucleo. -----A-----GC	(SEQ ID NO: 78)
	HUMCNPG2 green cone photo. -----T-AA	(SEQ ID NO: 79)
	HUMCNPR2 green cone photo. -----T-AA	(SEQ ID NO: 80)
35	HUMPIGMF2 colour-blind pho. -----T-AA	(SEQ ID NO: 81)

-72-

Table 22. c-jun specific probe (C2147).

	Subtype	GenBank name	C2147.DNA CCACGGCCAACATGCTC	
5	Human			
	B:	HUMJUNCA	-G---A---C---CAGC	(SEQ ID NO: 82)
	C:	HUMJUNA	-----	(SEQ ID NO: 83)
	D:	HUMJUNDR	-----G-G-C---G	(SEQ ID NO: 84)
10	Mouse			
	B:	MUSJUNBA	-G-----C---CAGC	(SEQ ID NO: 85)
	C:	MUSJUN	-----	(SEQ ID NO: 87)
		MUSCJUN	-----	(SEQ ID NO: 88)
		MUSJUNC	-----	(SEQ ID NO: 89)
15	D:	MUSJUND	---C---G-C---G	(SEQ ID NO: 90)
		MUSJUNDR	---C---G-C---G	(SEQ ID NO: 91)
		MUSJUNDA	---C---G-C---G	(SEQ ID NO: 92)
	Rat			
	C:	RATJUNAP	-----	(SEQ ID NO: 93)
20		RATRJG9	---C-----	(SEQ ID NO: 94)
	Chicken			
	C:	CHKJUN	-----	(SEQ ID NO: 95)
	Quail			
	C:	QULJUN	---T-----	(SEQ ID NO: 96)
25	Drosophila			
	C:	DROJUN	CCTACA-----ACC	(SEQ ID NO: 97)
	Homologous sequences in GenBank			
	MXBPALPA	L.enzymogenes	-----CG--	(SEQ ID NO: 98)
30	MXBPALP	L.enzymogenes	-----CG--	(SEQ ID NO: 99)

-73-

Table 23. Human jun-D specific probe (HUMD965)

	Subtype	GenBank name	HUMD965 ACACGCAGGAGCGCATC	
5	Human			
	B:	HUMJUNCA	GG-A--T-----	(SEQ ID NO: 100)
	C:	HUMJUNA	-GT-C-----G--	(SEQ ID NO: 101)
	D:	HUMJUNDR	-----	(SEQ ID NO: 102)
10	Mouse			
	B:	MUSJUNBA	-AGAC-----	(SEQ ID NO: 103)
	C:	MUSJUNC	-GT-T-----G--	(SEQ ID NO: 104)
		MUSCJUN	-GT-T-----G--	(SEQ ID NO: 105)
		MUSJUN	-GT-T-----G--	(SEQ ID NO: 106)
15	D:	MUSJUND	-----A--A-----	(SEQ ID NO: 107)
		MUSJUNDA	-----A--A-----	(SEQ ID NO: 108)
		MUSJUNDR	-----A--A-----	(SEQ ID NO: 109)
	Rat			
20	C:	RATJUNAP	-GT-T-----G--	(SEQ ID NO: 110)
		RATRJG9	-GT-T-----G--	(SEQ ID NO: 111)
	Chicken			
	C:	CHKJUN	-GT-----A-A--	(SEQ ID NO: 112)
	Quail			
25	C:	QULJUN	-GT-----A-A--	(SEQ ID NO: 113)
	Drosophila			
	C:	DROJUN	G--A--T-----	(SEQ ID NO: 114)
	Homologous sequences in GenBank			
	CELPOL11	C.elegans RNA	-----A-AA	(SEQ ID NO: 115)
	ECOPRIAY	E.coli primo.	-ACA-----	(SEQ ID NO: 116)
30	SINOCK82	Ockelbo 82 g. C	-----TGC	(SEQ ID NO: 117)
	TRPPROC	Treponema pa.	G-----TGC	(SEQ ID NO: 118)
	ECOCYSJ1HA	E.coli NADPH-	-----T-----G-G	(SEQ ID NO: 119)



-74-

Table 24. Mouse jun-D specific probe (MUSD1063).

	Subtype	GenBank name	MUSD1063.DNA CCCTCAAAAGCCAGAACACCG	
5	<hr/>			
	Human			
	B:	HUMJUNCA	-G-----GGC-G-----G-G-	(SEQ ID NO: 120)
	C:	HUMJUNA	AA--GC-C-----GC	(SEQ ID NO: 121)
	D:	HUMJUNDR	-----G--T-----G-	(SEQ ID NO: 122)
10	Mouse			
	B:	MUSJUNBA	TA---TCC---TCTG---C	(SEQ ID NO: 123)
	C:	MUSJUNC	AA--GC-T-----GC	(SEQ ID NO: 124)
		MUSJUN	AA--GC-T-----GC	(SEQ ID NO: 125)
15		MUSCJUN	AA--GC-T-----GC	(SEQ ID NO: 126)
	D:	MUSJUNDA	-----	(SEQ ID NO: 127)
		MUSJUNDR	-----	(SEQ ID NO: 128)
		MUSJUND	-----	(SEQ ID NO: 129)
20	Rat			
	C:	RATRJC9	G-----GCC---GC-C---TT	(SEQ ID NO: 130)
		RATJUNAP	G-----GCC---GC-C---TT	(SEQ ID NO: 131)
	Chicken			
	C:	CHKJUN	A---G-GG-A-A-----G---	(SEQ ID NO: 132)
25	Quail			
	C:	QULJUN	---C--G-T-----TG--G-A	(SEQ ID NO: 133)
	Drosophila			
	C:	DROJUN	A-----GGA---TGTGGCGC	(SEQ ID NO: 134)
30	Homologous sequences in GenBank			
	M27221		TG-----T--T-----	(SEQ ID NO: 135)
	DROAMY	D.erecta	G-----G-A--T-----	(SEQ ID NO: 136)
	DROAMYQ	D.erecta	G-----G-A--T-----	(SEQ ID NO: 137)
	HUMIGCMUDE	Immunoglob.	A--C-----GTT	(SEQ ID NO: 138)

-75-

Table 25. Human-rodent common jun-B specific probes.

	Name in GenBank	B-504 (5'-3') CACGACTACAACTCCTGAAAC	Seq. ID No:	B-739 (5'-3') GGACAGTACTTTTACCCCGG	Seq. ID No:	Size (bp)
House						
+	jun-B 5	MUSJUNBA	600		321	251
	c-JUN	MUSCJUN	601	ACC----T---G-G----AA	322	
o	jun-D	MUSJUND	602	ACG----T-C-C-----GAA	323	
					324	
Human						
	jun-B	HUMJUNCA	603		325	251
	c-jun 10	HUMJUNA	604	---A--C-GCGG-----AC	326	
	jun-D	HUMJUNDR	605	TC-----T-C-C-----AA	327	
GenBank						
	MUSMET1	-----T---CG-----GTA	606			
	RATHYHCD2			-----C-----AT---A	328	

15

\*: Size of PCR products using B-504 and B-739 as primers.

\*\*: The highest matched sequences in GenBank (release 68.0) next to the jun genes.

-76-

Table 26. Human-rodent common c-jun specific probes.

	Name in GenBank	C-2101 (5'-3') GAGGAAAAGTGAAAACCTTGAAAGC	Seq. ID No:	C-2219 (5'-3') GCCAACTCATGCTAACGCAG	Seq. ID No:	Size (bp)
Mouse						
jun-B 5	MUSJUNBA	-----C--G-----G--AC-C--G--	614	----GT-GC-----GG-GTC	329	
c-JUN	MUSCJUN	-----	615	-----	330	138
jun-D	MUSJUND	-----G-----C--G---C-C---AG	616	T----A-GCC--A-G----A	331	
Human						
jun-B	HUMJUNCA	-----C--G-----G--GC-C--G--	617	----G-C-CC-AG-----GC	332	
c-jun 10	HUMJUNA	-----	618	-----	333	138
jun-D	HUMJUNDR	--A--G-----G---C-C--GAG	619	T----G-GCC--A-G----A	334	
GenBank						
	HUMGASTA	A-A-----AA-----A---A-----	620			
	MUSOCT22			C-----G---GGC	335	

15

\*: Size of PCR products using C-2101 and C-2219 as primers.

\*\*: The highest matched sequences in GenBank (release 68.0) next to the jun genes.

-77-

Table 27. Human-rodent common jun-D specific probes.

	Name in GenBank	D-916 (5'-3') GACGTGCCGAGCTTCGG	Seq. ID No:	D-1153 (5'-3') AAAGTCCTCAGCCACGTCAAC	Seq. ID No:	Size (bp)
Mouse						
jun-B 5	MUSJUNBA	CTG-G-----ACT--	628	--G---A-G-C---T----G-	336	258
c-JUN	MUSCJUN	T-A-----AG---AGAC	629	-----A-G-A-----T---	337	
jun-D	MUSJUND	-----	630	-----	338	
Human						
jun-B	HUMJUNCA	C---C---CC-----C	631	--G---A-G-C-----G-	339	258
C-jun 10	HUMJUNA	T-A---TT-----CG--	632	-----A-G-A-----T---	340	
jun-D	HUMJUNDR	-----	633	-----	341	
GenBank						
	HUMTRKR	-----TG-----	634			
	HUMBSSL			---A-----A---CTG	342	

15

\*: Size of PCR products using D-916 and D-1153 as primers.

\*\*: The highest matched sequences in GenBank (release 68.0) next to the jun genes.

### Identifying Probes and PCR Primers

Probes and PCR primers for use in the methods of the present invention can be identified in any way known to the art. Preferably, however, such probes and primers are identified by a computer. We have developed a novel computer system for identifying the sequences to be used in such probes and primers. This system is an automated system which allows the user to calculate and design extremely accurate oligonucleotide probes and PCR primers.

The software of the present invention runs under Microsoft Windows® on IBM® compatible personal computers (PC's). This invention allows a researcher to design oligonucleotide probes based on the GenBank database of DNA and mRNA sequences. The present invention further allows examination of probes for specificity or commonality with respect to a user-selected target gene sequences. Hybridization strength between a probe and a target subsequence of DNA or mRNA can be estimated through a hybridization strength model. Quantitatively, hybridization strength is given as the melting temperature ( $T_m$ ).

Two models for estimating hybridization strength models are supported by this invention: 1) the Mismatch Model and 2) the H-Site Model. In either case, the user can select the following calculations for each probe, results of which are then made available for display and analysis: 1) Sequence, Melting Temperature ( $T_m$ ) and Hairpin characteristics (a hairpin is a nucleotide sequence that is homologous to itself and can "fold back" with one portion of the probe hybridizing to another portion of the same probe); 2) Hybridization to other species within the preparation mixture; and (3) Location and  $T_m$  for the strongest hybridizations. The results of the invention's calculations are then displayed on a Mitsunashi Probe Selection Diagram (MPSD) which is a graphic display of all potential hybridizations between the target mRNA and the probe sequences in the preparation.

The Main OligoProbe Design Station dialog window controls all user-definable settings in the program. The user is offered a number of options at this window. The File option allows the user to print, print in color, save selected probes, and exit the program. The Preparation option allows the user to open and create preparation (PRP) files. The Models option allows the user to choose between the two hybridization models currently supported by the OligoProbe DesignStation: 1) the H-Site Model and 2) the Mismatch Model.

If the user selects the H-Site Model option, the melting temperature for each probe and the nucleation threshold parameters can be set. The nucleation threshold is the number of base pairs constituting a nucleation site (a subsequence with an exact match). If the user selects the Mismatch Model option, the probe length and mismatches (N) can be set.

#### Mismatch Model

The Mismatch Model is used for designing DNA and mRNA probes utilizing sequence database information from sources such as GenBank. In this Model, hybridization strength is related only to the number of base pair mismatches between a probe and its target. Generally, the more mismatches a user allows when setting parameters, the more probes will be identified. The Mismatch Model does not take into account the GC content of candidate probes so there is no calculation of the probe's binding strength.

The basic technologies employed by the Mismatch model are hashing and continuous seed filtration. Hashing involves the application of an algorithm to the records in a set of data to obtain a symmetric grouping of the records. When using an indexed set of data such as a database, hashing is the process of transforming a record key to an index value for storing and retrieving a record. The Mismatch Model is essentially a quick process for determining exact and inexact matching between DNA and mRNA sequences to support the Mitsuhashi Probe Selection Diagram (MPSD).

The algorithm used by the Mismatch Model is based on the Waterman-Pevzner Algorithm (WPALG), which is a computer-based probe selection process. Essentially, this is a combination of new and improved pattern matching processes. See Hume and Sunday (1991, Ref. 4), Landau *et al* (1986-1990, Refs. 6, 7, 8), Grossi and Luccio (1989, Ref. 3), and Ukkonen (1982, Ref. 14).

There are three principal programs that make up the Mismatch Model in this implementation of the invention. The first is designated by the inventors as "k\_diff." WPALG uses k\_diff to find all locations of matches of length greater than or equal to one (1) (length is user-specified) with less than or equal to k number of mismatches (k is also user-specified) between the two sequences. If a candidate oligonucleotide probe fails to match that well, it is considered unique. k\_diff uses hashing and continuous seed filtration, and looks for homologs by searching GenBank and other databases with similar file formats. The technique of continuous seed filtration allows for much more efficient searching than previously implemented techniques.

-80-

A seed is defined in this invention to be a subsequence having a length equal to the longest exact match in the worst case scenario. For example, suppose the user selects a probe length (l) of 18, with 2 or fewer mismatches (k). If a match exists with 2 mismatches, then there must be a perfectly matching subsequence of length equal to 6.

5 Once the seed length has been determined, the Mismatch Model looks at all substrings of that seed length (in this example, the seed length would be 6), finds the perfectly matched base pair subsequence of length equals 6, and then looks to see if this subsequence extends to a sequence of length equal to the user selected probe length (i.e., 18 in this example). If so, a candidate probe has been found that meets the user's criteria.

10 Where the seed size is large (i.e., a long string of unique nucleotides), the program allocates a relatively large amount of memory for the hash table. This invention has an option that allows memory allocation for GenBank entries just once at the beginning of the program, instead of reallocating memory for each GenBank entry. This reduces input time for GenBank entries by as much as a factor of two (2), but this method requires the

15 user to know the maximum GenBank entry size in advance.

A probe is found to hybridize if it has k or fewer mismatches with a target sequence from the database or file searched. The hit extension time for all appropriate parameters of the Mismatch Model has been found by experimentation to be less than thirty-five (35) seconds, except in one case where the minimum probe length (l) was set to 24 and the

20 maximum number of mismatches (k) was set to four (4). This situation would rarely be used in real gene localization experiments because the hybridization conditions are too weak.

#### H-Site Model

In this embodiment of the invention, the second hybridization strength model is termed the H-Site Model. One aspect of the H-Site Model uses a generalization of an experimental formula to analyze nucleotide binding strength. The basic formula on which this aspect of the model is built is as follows:

25

$$T_m = 81.5 - 16.6(\log[Na]) - .63 \%(\text{formamide}) + .41 (\%(\text{G} + \text{C})) - 600 / N$$

In this formula, log[Na] is the log of the sodium concentration,  $\%(\text{G} + \text{C})$  is the fraction of matched base pairs which are G-C complementary, and N is the probe length. This formula relates the fact that melting temperature is a function of both probe length and percent GC content. This basic formula has been modified in this invention to account for the presence of mismatches. Each percent of mismatch reduces the melting temperature by an average of 1.25° (2°C for an AT mismatch, and 4°C for a GC

30

mismatch). This formula is, however, an approximation. The actual melting temperature might potentially differ from this approximation, especially for short probes or probes with a relatively large number of mismatches.

Hybridization strength in the H-Site Model is related to each of the following factors:  
5 1) "binding region": 2) type of mismatch (GC or AT substitution); 3) length of the probe;  
4) GC content of the binding region; and 5) existence of a "nucleation site" (a subsequence with an exact match). The type of mismatch and GC content of the binding region from each sequence contributes to a candidate probe's binding strength. The binding strength from each probe is thereby determined enabling the user to select an optimal probe.

10 The fundamental assumption of the H-Site Model is that binding strength is mostly determined by a paired subsequence of the probe and target, called the binding region. If the subsequence binding region contains more GC pairs than AT pairs, the binding strength will be higher due to the greater number of hydrogen bonds between G and C bases (three bonds) in comparison to A and T bases (two bonds). Thus, GC rich probes  
15 have a higher melting temperature and subsequently form stronger hybridizations.

In the H-Site Model the program determines optimal probes, ideally without any mismatches to the target gene. With this model, however, a candidate probe can have more AT mismatches if the sequence is GC rich. The amount of allowable AT mismatches in a specific sequence is determined in the present invention program by looking primarily  
20 at subsequence regions of the probe and target that match without penalizing the probe for areas that mismatch. If the mismatches are located at either or both of the ends of the binding region, there is little effect on the overall stability of the base-pairing. Centrally located mismatches in the binding region are much more deleterious, as this will significantly lower the binding strength of the probe.

25 The formula cited above for the melting temperature applies within the binding region. The length of the probe is used to calculate percentages, but all other parameters of the formula are applied to the binding region only. The H-Site Model further assumes the existence of a nucleation site. The length of this nucleation site may be set by the user. Typically, a value of 8 to 10 base pairs is used. To complete the H-Site Model, the  
30 binding region is chosen so as to maximize the melting temperature  $T_m$  among all regions containing a nucleation site, assuming one exists (otherwise,  $T_m = 0$ ).

The H-Site Model is more complex than the Mismatch Model discussed above in that hybridization strength is modeled as a sum of multiple subsequence contributions, with matches generally providing positive binding energy and mismatches generally providing



negative binding energy. The exact binding energies to be used depend only on the matched or mismatched pair. These coefficients may be specified by the user, although in the current version of this invention these coefficients are not explicitly user-selectable, but rather are selected to best fit the hybridization strength formulas developed by Itakura et al (1984, Ref. 5), Bolton and McCarthy (1962, Ref. 2), Benner et al (1973, Ref. 1), and Southern (1975, Ref. 13).

A unique aspect of the H-Site Model is that hybridization strength is determined by the optimal binding region between the candidate probe and binding locus. This binding region is called the hybridization site, or h-site, and is selected so as to maximize overall hybridization strength, so that mismatches outside the binding region do not detract from the estimated hybridization strength. Several other unique features of the H-Site Model include the fact that it is more oriented toward RNA and especially cDNA sequences than DNA sequences, and the fact that the user has control over preparation and environmental variables.

The emphasis on RNA and cDNA sequences allows the user to concentrate on coding regions of genes, rather than necessitating sorting through all of a genomic sequence for the desired probe. The enhanced user control over environmental and preparation variables allows the user to more accurately simulate laboratory conditions that closely correspond with any experiments he or she is conducting. Further, this implementation of the invention does some preliminary preprocessing of the GenBank database to sort out and select the cDNA sequences. This is done by locating a keyword (in this case CDS) in each GenBank record, thereby eliminating any sequences containing introns.

The Mitsuhashi Probe Selection Diagram (MPSD), Figure 28, is a key feature of this invention, as it is a unique way of visualizing the results of the probe designed by the Mismatch and H-Site Models. It is a graphic display of all of the hybridizations of candidate oligonucleotide probes and the target with all sequences in the preparation. Given a gene sequence database and a target mRNA sequence, the MPSD graphically displays all of the candidate probes and their hybridization strengths with all sequences from the database. In the present implementation, each melting temperature is displayed as a different color, from red (highest  $T_m$ ) to blue (lowest  $T_m$ ). The MPSD allows the user to see visually the number of false hybridizations at various temperatures for all candidate probes, and the sources of these false hybridizations (with a loci and sequence comparison). A locus may be a specific site or place, or, in the genetic sense, a locus is

any of the homologous parts of a pair of chromosomes that may be occupied by allelic genes.

These probes may then be used to test for the presence of precursors of specific proteins in living tissues. The oligonucleotide probes designed with this invention may be used for medical diagnostic kits, DNA identification, and potentially continuous monitoring of metabolic processes in human beings. The present implementation of this computerized design tool runs under Microsoft® Windows™ v. 3.1 (made by Microsoft Corporation; Redmond, Washington) on IBM® compatible personal computers (PC's).

The H-Site Model of this invention is unique in that it offers a multitude of information on selected probes and original and distinctive means of visualizing, analyzing and selecting among candidate probes designed with the invention. Candidate probes are analyzed using the H-Site Model for their binding specificity relative to some known set of mRNA or DNA sequences, collected in a database such as the GenBank database. The first step involves selection of candidate probes at some or all the positions along a given target. Next, a melting temperature model is selected, and an accounting is made of how many false hybridizations each candidate probe will produce and what the melting temperature of each will be. Lastly, the results are presented to the researcher along with a unique set of tools for visualizing, analyzing and selecting among the candidate probes.

This invention is both much faster and much more accurate than the methods that are currently in use. It is unique because it is the only method that can find not only the most specific and unique sequence, but also the common sequences. Further, it allows the user to perform many types of analysis on the candidate probes, in addition to comparing those probes in various ways to the target sequences and to each other.

Therefore, it is the object of this invention to provide a practical and user-friendly system that allows a researcher to design both specific and common oligonucleotide probes, and to do this in less time and with much more accuracy than currently done.

This invention is employed in the form best seen in Figure 20. There, the combination of this invention consists of an IBM® compatible personal computer (PC), running software specific to this invention, and having access to a distributed database with the file formats found in the GenBank database and other related databases.

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred

methods and materials are now described. All publications mentioned hereunder are incorporated herein by reference.

The preferred computer hardware capable of operating this invention involves of a system with at least the following specifications (Figure 20): 1) an IBM® compatible PC, generally designated 1A, 1B, and 1C, with an 80486 coprocessor, running at 33 Mhz or faster; 2) 8 or more MB of RAM, 1A; 3) a hard disk 1B with at least 200 MB of storage space, but preferably 1 GB; 4) a VGA color monitor (1C) with graphics capabilities of a size sufficient to display the invention's output in readable format, preferably with a resolution of 1024 x 768; and 5) a 580 MB CD ROM drive 5 (1B of Figure 20 generally refers to the internal storage systems included in this PC, clockwise from upper right, two floppy drives, and a hard disk). Because the software of this invention preferably has a Microsoft® Windows™ interface, the user will also need a mouse 2, or some other type of pointing device.

The preferred embodiment of this invention would also include a laser printer 3 and/or a color plotter 4. The invention may also require a modem (which can be internal or external) if the user does not have access to the CD ROM versions of the GenBank database 8 (containing a variable number of gene sequences 6). If a modem is used, information and instructions are transmitted via telephone lines to and from the GenBank database 8. If a CD ROM drive 5 is used, the GenBank database (or specific portions of it) is stored on a number of CDs.

The computer system should preferably have at least the Microsoft® DOS version 5.0 operating system running Microsoft® Windows™ version 3.1. All of the programs in the preferred embodiment of the invention were written in the Borland® C++ (Borland International, Inc.; Scotts Valley, CA) computer language. It should be noted that subsequently developed computers, storage systems, and languages may be adapted to utilize this invention and vice versa.

This inventive computer program is designed to enable the user to access DNA, mRNA and cDNA sequences stored either in the GenBank or in databases with similar file formats. GenBank is a distributed flat file database made up of records, each record containing a variable number of fields in ASCII file format. The stored database itself is distributed, and there is no one database management system (DBMS) common to even a majority of its users. One general format, called the line type format, is used both for the distributed database and for all of GenBank's internal record keeping. All data and system files and indexes for GenBank are kept in text files in this line type format.

The primary GenBank database is currently distributed in a multitude of files or divisions, each of which represents the genome of a particular species (or at least as much of it as is currently known and sequenced and publicly available). The GenBank provides a collection of nucleotide sequences as well as relevant bibliographic and biological annotation. Release 72.0 (6/92) of the GenBank CD distribution contains over 71,000 loci with a total of over ninety-two (92) million nucleotides. GenBank is distributed by IntelliGenetics, of Mountain View, CA, in cooperation with the National Center for Biotechnology Information, National Library of Medicine, in Bethesda, MD.

10 1. Overall Description of the OligoProbe DesignStation

a. **General Theory**

The intent of this invention is to provide one or more fast processes for performing exact and inexact matching between DNA sequences to support the Mitsuhashi Probe Selection Diagram (MPSD) discussed below, and other analysis with interactive graphical analysis tools. Hybridization strength between a candidate oligonucleotide probe and a target subsequence of DNA, mRNA or cDNA can be estimated through a hybridization strength model. Quantitatively, hybridization strength is given as the melting temperature ( $T_m$ ). Currently, two hybridization strength models are supported by the invention: 1) the Mismatch Model and 2) the H-Site Model.

20 b. **Inputs**

i. Main OligoProbe DesignStation Dialog Window

The Main OligoProbe DesignStation dialog window, Figure 21, controls all user-definable settings. This window has a menu bar offering five options: 1) File 10; 2) Preparation 80; 3) Models 70; 4) Experiment 90; and 5) Help 50. The File 10 option allows the user to print, print in color, save selected probes, and exit the program. The Preparation 80 option allows the user to open and create preparation (PRP) files.

The Models 70 option allows the user to choose between the two hybridization models currently supported by the OligoProbe DesignStation: 1) the H-Site Model 71 and 2) the Mismatch Model 75. If the user selects the H-Site Model 71 option, the left hand menu of Figure 21C is displayed and the user sets the following model parameters: 1) the melting temperature  $T_m$  72 for which probes are being designed (i.e., the melting temperature that corresponds to a particular experiment or condition the user desires to simulate); and 2) the nucleation threshold 73, which is the number of base pairs constituting a nucleation site. If the user selects the Mismatch Model 75 option, the right

hand menu of Figure 21C is displayed and the user sets the following model parameters:  
1) probe length 76, which is the number of base pairs in probes to be considered; and 2)  
mismatch N 77, which is the maximum number of mismatches constituting a hybridization.  
Computation of the user's request takes longer with the H-Site Model if the threshold 73  
5 setting is decreased, but longer with the Mismatch Model if the number of mismatches K  
77 is increased.

In addition, for both Model options the user chooses the target species 11 DNA or  
mRNA for which probes are being designed and the preparation 12, a file of all sequences  
with which hybridizations are to be calculated. A sample of a target species file is shown  
10 in Figure 47 (humbjunix.cds), while a sample of a preparation file is shown in Figure 48  
(junmix.seq). Each of these inputs is represented by a file name and extension in standard  
DOS format. In the target species and preparation fields, the file format follows the  
GenBank format with each of the fields having a default file extension. Pressing the "OK"  
button 91 (Figure 21C) will initiate processing, while pressing the "Cancel" button 93 will  
15 stop the processing.

The Experiment 90 option and the Help 50 option are expansion options not yet  
available in the current implementation of the invention.

#### c. Processing

Figure 22 is a flow chart of the overall OligoProbe DesignStation Program, illustrating  
20 its sequence and structure. Generally, the main or "control" program of the OligoProbe  
DesignStation performs overall maintenance and control functions. This program, as  
illustrated in Figure 22, accomplishes the general housekeeping functions 51, such as  
defining global variables. The user-friendly interface 53, carries out the user-input  
procedures 55, the file 57 or database 59 access procedures, calling of the model program  
25 62 or 63 selected by the user, and the user-selected report 65 or display 67, 69, 71 and 73  
features. Each of these features is discussed in more detail in later sections, with the  
exception of the input procedures, which involves capturing the user's set-up and control  
inputs.

#### d. Outputs

##### i. The Mitsuhashi Probe Selection Diagram Window

30 The Mitsuhashi Probe Selection Diagram (MPSD), Figure 23, is a key feature of the  
invention as it is a unique way of visualizing the results of the program's calculations. It  
is a graphic display of all of the hybridizations of probes with the target oligonucleotides  
in the preparation. Specifically, given a nucleotide sequence database and a target mRNA,

the MPSD graphically displays all of the candidate probes and their hybridization strengths with all sequences from the nucleotide database. The MPSD allows the user to visualize the number of false hybridizations at various temperatures for all candidate probes, and the sources of these false hybridizations (with a loci and sequence comparison).

5 For each melting temperature selected, a graph showing the number of hybridizations for each probe is displayed. In the preferred embodiment, the graphs are color coded. In this implementation of the invention, the color red 123 identifies the highest melting temperature and the color blue 124 identifies the lowest melting temperature. Each mismatch results in a reduction in the  $T_m$  value. The melting temperature is also a  
10 function of probe length and percent GC content. Within the window, the cursor 125 shape is changed from a vertical line bisecting the screen to a small rectangle when the user selects a particular probe. The current probe is defined to be that probe under the cursor position (whether it be a line or a rectangle) in the MPSD window. More detailed information about the current probe is given in the ProbeInfo and MatchInfo windows,  
15 discussed below. Clicking the mouse button 2 once at the cursor 125 selects the current probe. Clicking the mouse button 2 a second time deselects the current probe. Moving the cursor across the screen causes the display to change and reflect the candidate probe under the current cursor position.

The x-axis 110 of the MPSD, Figure 23, shows the candidate probes' starting positions  
20 along the given mRNA sequence. The user may "slide" the display to the left or right in order to display other probe starting positions. The y-axis 115 of the MPSD displays the probe specificity, which is calculated by the program.

The menu options 116, 117, 118, 119, and 120 available to the user while in the MPSD, Figure 23, and are displayed along a menu bar at the top of the screen. The user  
25 can click the mouse 2 on the preferred option to briefly display the option choices, or can click and hold the mouse button on the option to allow an option to be selected. The user may also type a combination of keystrokes in order to display an option in accordance with well-known computer desk top interface operations. This combination usually involves  
> holding down the ALT key while pressing the key representing the first letter of the  
30 desired option (i.e. F, P, M, E or H).

The File option 116 allows the user to specify input files and databases. The Preparation option 117 allows the user to create a preparation file summarizing the sequence database. The Models option 118 allows the user to specify the hybridization model (i.e., H-Site or Mismatch) and its parameters. The Experiment option 119 and the

Help option 120 are not available in the current implementation of this invention. These options are part of the original Main OligoProbe DesignStation dialog window, Figure 21.

Areas on the graphical display of the MPSD, Figure 23, where the hybridizations for the optimal probes are displayed are lowest and most similar, such as shown at 121. Areas on the graphical display of the MPSD where the hybridizations for the optimal probes are displayed are highest and most dissimilar, such as shown at 122, indicate that the particular sequence displayed is extremely specific to that particular gene fragment. The high points on the MPSD show many loci in the database, to which the candidate probe will hybridize (i.e., many false hybridizations). The low points show few hybridizations, at least relative to the given database. Specifically, the sequence shown at 121 would reflect a probe common to all of the gene fragments tested, such that this probe could be used to detect each of these genes. The sequence shown at 122 would reflect a probe specific to the particular gene fragment, such that this probe could be used to detect this particular gene and no others.

ii. The ProbeInfo and MatchInfo Window

The combined ProbeInfo and MatchInfo Window, Figure 24, displays detailed information about the current candidate probe. The upper portion of the window is the ProbeInfo window, and the lower portion is the MatchInfo window. The ProbeInfo window portion displays the following types of information: the target locus (i.e., the mRNA, cDNA, or DNA from which the user is looking for probes) is displayed at 131, while the preparation used for hybridizations is displayed at 132. In the example shown in Figure 24, the target locus 131 is the file named HUMBJUNX.CDS, which is shown as being located on drive F in the subdirectory MILAN. The preparation 132 is shown as being the file designated JUNMIX.PRP, which is also shown as being located on drive F in the subdirectory MILAN. The JUNMIX.PRP preparation in this example is a mixture of human and mouse jun loci.

The current and optimal probe's starting position is shown at 135. The current candidate oligonucleotide probe is defined at 136, and is listed at 137 as having a length of 21 bases. The melting temperature for the probe 136 as hybridized with the targets is shown in column 140. The melting temperature for the optimal probe is given as 61.7 degrees C at 138. The ProbeInfo Window Figure 24 also displays hairpin characteristics of the probe at 139. In the example shown, the ProbeInfo Window shows that there are

four (4) base pairs involved in the worst hairpin, and that the worst hairpin has a length of one (1) (see Figure 24, at 139).

The MatchInfo Window portion displays a list of hybridizations between the current probe and species within the preparation file, including hybridization loci and hybridization temperatures. The hybridizations are listed in descending order by melting temperature. The display shows the locus with which the hybridization occurs, the position within the locus, and the hybridization sequence.

In the MatchInfo window portion, the candidate probe 136 is shown at 150 as hybridizing completely with a high binding strength. This is because the target DNA is itself represented in the database in this case, so the candidate probe is seen at 150 to hybridize with itself (a perfect hybridization). The locus of each hybridization from the preparation 132 are displayed in column 141, while the starting position of each hybridization is given in column 142. The calculated hybridizations are shown at 145.

### iii. The ProbesEdit Window

The ProbesEdit Window, Figure 25, is a text editing window provided for convenient editing and annotation of OligoProbe DesignStation text file output. It is also used to accumulate probes selected from the MPSD, Figure 23, by mouse button 2 clicks. Standard text editing capabilities are available within the ProbesEdit Window. The user may accumulate selected probes in this window (see 155 for an example) and then save them to a file (which will bear the name of the preparation sequence with the file extension of ".prb" 156, or may be another file name selected by the user). A sample of this file is shown in Figure 26A.

### iv. Miscellaneous Output

The present embodiment of this invention also creates two output files, currently named "test.out" and "test1.out", depending upon which model the user has selected. The first file, "test.out", is created with both the Mismatch Model and the H-Site Model. This file is a textual representation of the Mitsuhashi Probe Selection Diagram (MPSD). It breaks the probe sequence down by position, length, delta Tm, screensN, and the actual probe sequence (i.e., nucleotides). An example of this file created by the Mismatch Model is shown in Figure 40, and example created by the H-Site Model is shown in Figure 44a. The second file, "test1.out", is created only by the H-Site Model. This file is a textual representation of the ProbeInfo and MatchInfo window that captures all hybridizations, along with their locus, starting position, melting temperature, and possible other



-9(-

hybridizations. A partial example of this file is shown in Figure 44b (10 pages out of a total of 190 pages created by the H-Site Model).

## 2. Description of the Mismatch Model Program

### a. **Overview**

In this invention, one of the hybridization strength models is termed the Mismatch Model (see Figure 21 for selection of this model). The basic operation of this model involves the techniques of hashing and continuous seed filtration, as defined earlier but described in more detail below. The essence of the Mismatch Model is a fast process for doing exact and inexact matching between nucleotide sequences to support the Mitsunashi Probe Selection Diagram (MPSD). There are a number of modules in the present implementation of the Mismatch Model contained in this invention, the most significant of which are shown in the flow chart in Figure 27 and in more detail in Figures 28 through 38. The main k\_diff module shown in the flow chart in Figure 28 is a structured program that provides overall control of the Mismatch Model, calling various submodules that perform different functions.

### b. **Inputs**

The user-selected input variables for this model are minimum probe length 76 (which is generally from 18 to 30) and maximum number of mismatches 77 (which generally is from 1 to 5). These inputs are entered by the user in the Main OligoProbe DesignStation Dialog Window, Figure 21C.

### c. **Processing**

#### i. k\_diff Program

Some terms of art need to be defined before the processing performed by this module can be explained. A hash table basically is an array or table of data. A linked list is a classical data structure which is a chain of linked entries and involves pointers to other entry structures. Entries in a linked list do not have to be stored sequentially in memory, as is the case with elements contained in an array. Usually there is a pointer to the list associated with the list, which is often initially set to point to the start of the list. A pointer to a list is useful for sequencing through the entries in the list. A null pointer (i.e., a pointer with a value of zero) is used to mark the end of the list.

As the flow charts in Figures 27 and 28 illustrate, the general process steps and implemented functions of this model can be outlined as follows:

-91-

Step 1: First, create a hash table and linked list from the query (Figure 27, hashing module 222).

Step 2: Next, while there are still GenBank entries available for searching (Figure 27, assembly module 230):

5

Step 2a: Read the current GenBank entry (record) sequence of user-specified length (Figure 27, seqload module 232), or read the current sequence (record) from the file selected by the user (Figure 27, read1 module 234).

10

Step 2b: For the current sequence for each position of the sequence from the first position (or nucleotide) to the last position (or nucleotide) (incrementing the position number once each iteration of the loop) (Figure 27, q\_colour module 242).

Step 2c: set the variable dna\_hash equal to the hash of the current position of the current sequence (Figure 27, q\_colour module 242).

15

Step 2d: While not at the end of the linked list for dna\_hash (Figure 27, q\_colour module 242).

Step 2e: set the query\_pos equal to the current position of dna\_hash in the linked list (Figure 27, q\_colour module 242) and

20

Step 2f: Extend the hit with the coordinates (query\_pos, dna\_pos) (Figure 27, hit\_ext module 244).

Step 2g: If there exists a k\_mismatch in the current extended hit (Figure 27, colour module 246), then

Step 2h: print the current hit (Figure 27, q\_colour module 242), and repeat from Step 2.

25

As this illustrates, there are three (3) basic looping or iteration processes with functions being performed based on variables such as whether the GenBank section end has been reached (the first "WHILE" loop, Step 2), whether the end of the current DNA entry has been reached (the "FOR" loop, Step 2b), and whether the end of the dna\_hash linked list has been reached (the second "WHILE" loop, Step 2d). A "hit" will only be

30

printed if there are k\_mismatches in the current extended hit.

Figures 28 through 38 illustrate the functions of each of the modules of the present embodiment of this invention, all of which were generalized and summarized in the description above. Figure 28, which outlines the main "k\_diff" module, shows that this module is primarily a program organization and direction module, in addition to

performing routine "housekeeping" functions, such as defining the variables and hash tables 251, checking if the user-selected gene sequence file is open 252, extracting needed identification information from the GenBank 253, and ensuring valid user input 254. This module also performs a one-time allocation of memory for the gene sequences, and  
5 allocates memory for hit information, hashing, hybridization and frequency length profiles and output displays, 255 & 256. The "k\_diff" module also initializes or "zeros out" the hashing table, the linked hashing list and the various other variables 257 in preparation for the hashing function. In addition, this module forms the hash tables 258 and extracts a sequence and finds the sequence length 259.

10 One of the most important functions performed by the "k\_diff" module is to define the seed (or kernel or k\_tuple) size. This is done by setting the variable k\_tuple equal to  $(\text{min\_probe\_length} - \text{max\_mismatch\_}\#) / (\text{max\_mismatch\_}\# + 1)$  Figure 28 at 265. Next, if the remainder of the aforementioned process is not equal to zero 266, then the value of the variable k\_tuple is incremented by one 267. The resulting value is the size of the  
15 seed. The module then reads the query 268 and copies the LOCUS name 269 for identification purposes (a definition of the term locus is given earlier in the specification).

The "k\_diff" module Figure 28 also calls the "assembly" module 260, writes the results to a file 261a, plots the results 261b (discussed below), calculates the hairpin characteristics  
20 262 (i.e., the number of base pairs and the length of the worst hairpin) and the melting temperature (Tm) for each candidate probe 263, and saves the results to a file 264.

The screen graphs are plotted 261b by converting the result values to pixels, filling a pixel array and performing a binary search into the pixel array. Next, given the number of pixels per probe position and which function is of interest to the user (i.e., the three mismatch match numbers), the program interpolates the values at the value of  
25 (pixelsPerPositionN-1) and computes the array of pixel values for drawing the graph. These values are then plotted on the MPSD.

The "hashing" module, Figure 29, performs hashing of the query. In other words, it creates the hash table and linked list of query positions with the same hash. The variable  
30 has\_table[i] equals the position of the first occurrence of hash i in the query. If i does not appear in the query, hash\_table[i] is set to zero.

The "tran" module, Figure 30, is called by the "hashing" module 271, and performs the hashing of the sequence of k\_tuple (kernel or seed) size. If the k\_tuple exists (i.e., its length is greater than zero), the variable uns is set equal to  $\text{uns} \cdot \text{ALF} + p$  291. The variable p represents the digit returned by the "let\_dig" module Figure 31 that represents the

nucleotide being examined. ALF is a constant that is set by the program in this implementation to equal four. The query pointer is then incremented, while the size of k\_tuple (the seed) is decremented 292. This process is repeated until the sequence of k\_tuple has been entirely hashed. Then the "tran" module returns the variable  
5 current\_hash 293 to the "hashing" module Figure 29.

The "let\_dig" module, Figure 31, is called by the "tran" module 291, and transforms the nucleotides represented as the characters "A", "T", "U", "G" and "C" in the GenBank and the user's query into numeric digits for easier processing by the program. This module transforms "a" and "A" into "0" 301, "t", "T", "u" and "U" into "1" 302, "g" and "G" into "2" 303, and "c" and "C" into "3" 305. If the character to be transformed does not  
10 match any one of those listed above, the module returns "-1" 305. The "hashing" module, Figure 29, then calls the "update" module 272, Figure 32, which updates the hash with a sliding window (i.e., it forms a new hash after shifting the old hash by "1"). The remainder of old\_hash divided by power\_1 is calculated 311 (a modulus operation), the remainder is multiplied by ALF 312 (i.e., four), and then the digit representing the nucleotide is added  
15 to the result 313. The "update" module then returns the result 314 to the "hashing" module Figure 29.

If the current hash has already occurred in the query, the program searches for the end of the linked list for the current hash 273 and marks the end of the linked list for the  
20 current hash 274. If the current hash has not already occurred in the query, the program puts the hash into the hash table 275. The resulting hash table and linked list are then returned to the "k\_diff" module, Figure 28 at 258.

The "assembly" module, Figure 33, extracts sequences from the GenBank and performs hit locating and extending functions. This module is called by the "k\_diff" module  
25 Figure 28 at 260 if the user has chosen to use the database to locate matches. The output from the "assembly" module (Figure 33) tells the user that the section of the database searched contains E number of entries 321 of S summary length 322 with H number of hits 323. Further, the program tells the user that the number of considered l-tuples equals T 324. The entry head line is also printed 326.

30 The "seqload" module, Figure 34, is called by the "k\_diff" module Figure 28 at 259 once the query hash table and linked list have been formed by the "hashing" module Figure 29. The "seqload" module Figure 34 checks to see if the end of the GenBank file has been reached 327, and, if not, searches until a record is found with LOCUS in the head-line 328.

-94-

Next, the LOCUS name is extracted 329 for identification purposes, and the program searches for the ORIGIN field in the record 330.

5 The program then extracts the current sequence 331 from the GenBank and performs two passes on each sequence. The first is to determine the sequence length 332 and allocate memory for each sequence 333, and the second pass is to read the sequence into the allocated memory 334. Since the sequences being extracted can contain either DNA nucleotides or protein nucleotides, the "seqload" module can recognize the characters "A", "T", "U", "G", and "C". The bases "A", "T", "G" and "C" are used in DNA sequences, while the bases "A", "U", "G" and "C" are used in RNA and mRNA sequences. The extracted  
10 sequence is then positioned according to the type of nucleotides contained in the sequence 335, and the process is repeated. Once the end of the sequence has been reached, the "seqload" module returns the sequence length 336 to the "k\_diff" module Figure 28.

If the user has chosen to use one or more files to locate matches, rather than the database, the "read1" module, Figure 35, rather than the "seqload" module Figure 34, is  
15 called by the "k\_diff" module Figure 28. The "read1" module, Figure 35, reads the sequence from the user specified query file 341 and allocates memory 342. This module also determines the query length 343, extracts sequence identification information 344, determines the sequence length 345, transforms each nucleotide into a digit 346 by calling the "let\_dig" module Figure 31, creates the query hash table 347 by calling the "dig\_let" module Figure 36, and closes the file 348 once everything has been read in.  
20

First, the "read1" module Figure 35 allocates space for the query 342. To do this, the "ckalloc" module, Figure 35 at 342, is called. This module allocates space and checks whether this allocation is successful (i.e., is there enough memory or has the program run out of memory). After allocating space, the "read1" module Figure 35 opens the user-  
25 specified file 349 (the "ckopen" module, Figure 35 at 349, is called to ensure that the query file can be successfully opened 349), determines the query length 343, locates a record with LOCUS in the head-line and extracts the LOCUS name 344 for identification purposes, locates the ORIGIN field in the record and then reads the query sequence from the file 341. Next, the sequence length is determined 345, memory is allocated for the sequence  
30 342, and the sequence is read into the query file 350. If the string has previously been found, processing is returned to 344. If not, then each character in the query file is read into memory 350.

The characters are transformed into digits 346 using the "let\_dig" module, Figure 31, until a valid digit has been found, and then the hash table containing the query is set up

347 using the module "dig\_let". Figure 36, which transforms the digits into nucleotides represented by the characters "A" 371, "T" 371, "G" 373, "C" 374, and "X" 375 as a default. If the end of the file has not been reached, processing is returned to 344. If it has, the file is closed 348 and the query is then returned to the "read1" module Figure 35 at 347.

5       The "q\_colour" module, Figure 27 (Figure 33 at 325), is called by the "assembly" module Figure 33 after the current sequence has been extracted from the GenBank. The "q\_colour" module Figure 37 performs the heart of the Mismatch Model process in that it performs the comparison between the query and the database or file sequences. If the module finds that there exists a long (i.e., greater than the min\_hit\_length) extended hit,  
10       it returns a "1" to the "assembly" module Figure 34. Otherwise, the "q\_colour" module, Figure 37, returns a "0".

      In the "q\_colour" module, Figure 37, all DNA positions are analyzed in the following manner. First, the entire DNA sequence is analyzed 391 to see whether each position is equal to zero 392 (i.e., whether it is empty or the sequence is finished). If it is not equal  
15       to zero 393, the "q\_colour" module Figure 40 calls the "tran" module, Figure 30 described above, which performs the hashing of k\_tuples. The "tran" module Figure 30 calls other modules which transform the nucleotides represented by characters into digits for easier processing by the program and then updates the hash with a sliding window. If the position is equal to zero, the current\_hash position is set to new\_has after one shift of  
20       old\_hash 390 by calling the "update" module Figure 32.

      If the nucleotide at the current\_hash position is equal to zero, processing is returned to 391. If not, the query position is set equal to (nucleotide at current hash position - 1). Next, the "q\_colour" module Figure 37 looks for the current\_hash in the hash table 394. If the current k\_tuple does not match the query 395, then the next k\_tuple is considered  
25       395, and processing is returned to 391. If the current k\_tuple does match the query, then the program checks the hit's (i.e., the match's) vicinity 396 by calling the "hit\_ext" module, Figure 38 to determine if the hit is weak. The inventors have found that if the code for the module "hit\_ext" is included within the module "q\_colour", rather than being a separate module utilizing the parameter transfer machinery, 25% of CPU time can be saved.

30       The "hit\_ext" module Figure 38 determines the current query position in the hit's vicinity 421, determines the current DNA position in the hit's vicinity 422, and creates the list of mismatch positions (i.e., the mismatch\_location\_ahead 423, the mismatch\_location\_behind 423 and the kernel match location). If the hit is weak 424, the "hit\_ext" module Figure 38 returns "0" to the "q\_colour" module Figure 37. If the hit has

-96-

a chance to contain 425, the module returns "1" to the "q\_colour" module Figure 37. A hit has a chance to contain, and is therefore not considered weak, if the mismatch\_location\_ahead - the mismatch\_location\_behind is greater than the min\_hit\_length. If not, it is a short hit and is too weak.

5        If the "hit\_ext" module Figure 38 tells the "q\_colour" module Figure 37 that the hit was not a weak one, then the "q\_colour" module determines whether the current hit is long enough 398 by calling the "colour" module Figure 39. The "colour" module Figure 39 performs query\_colour modification by the hit data, starting at pos\_query and described by mismatch\_location\_ahead and mismatch\_location\_behind. After the variables to be used in this module are defined, variable isw\_print (which is the switch indicating the hit length) is initialized to zero 430. The cur\_length is then set equal to the length of the extending hit 431 (mismatch\_location\_behind[i] + mismatch\_location\_ahead[j]-1). Next, if cur\_length is greater than or equal to the min\_hit\_length 432 (i.e., the minimum considered probe size), the hit is considered long and isw\_print is set equal to two 433. 10        The value of isw\_print is then returned 434 to the "q\_colour" module Figure 37.

15        If the length of the extending hit is longer than the min\_hit\_length, the hit is considered long 399. Otherwise, the hit is considered short. If the hit is short, nothing more is done to the current hit and the module begins again. If, on the other hand, the hit is considered long 399, the "q\_colour" module Figure 37 prints the current extended hit 400. The current extended hit can be printed in ASCII, printed in a binary file, or printed to a memory file. The "q\_colour" module Figure 37 then repeats until the end of the linked list is reached.

d.            **Outputs**

25        The output of the k\_diff program may be either a binary file containing the number of extended hits and the k\_mismatch hit locations (see Figure 40), or the output may be kept in memory without writing it to a file. See Section I(d)(iv) for more detail.

3.    Description of the H-Site Model Program

a.            **Overview**

30        In this invention, the second hybridization strength model is termed the H-Site Model (see Figure 21 for user selection of this model). The formula used in the H-Site Model is an expression of the fact that melting temperature  $T_m$  is a function of both probe length and percent of GC content. This basic formula has been modified in this invention to account for the presence of mismatches. Each percent of mismatch reduces the melting

temperature  $T_m$  by an average of 1.25 degrees (2 degrees C for an AT mismatch, and 4 degrees C for a GC mismatch).

In addition, this implementation of the invention does some preliminary preprocessing of the GenBank database to sort out and select the cDNA sequences. This is done by  
5 locating a keyword (in this case CDS) in each GenBank record.

There are a number of modules in the present embodiment of the H-Site Model contained in this invention. Each step of the processing involved in the H-Site Model is more fully explained below, and is accompanied by detailed flow charts.

b. **Inputs**

10 There are two basic user-selected inputs for the H-Site Model (see Figure 21C): 1) the melting temperature  $T_m$  22 for which probes are being designed (i.e., the melting temperature that corresponds to a particular experiment or condition the user desires to simulate); and 2) the nucleation threshold 23, which is the number of base pairs constituting a nucleation site. The user is also required to select the 1) target species 11  
15 gene sequence(s) (DNA, mRNA or cDNA) for which probes are being designed; 2) the preparation 12 of all sequences with which hybridizations are to be calculated; and 3) the probe output file 13. The preparation file is the most important, as discussed below.

c. **Organization of the H-Site Model Program**

20 The current implementation of the H-Site Model program of this invention is distributed between five files containing numerous modules. The main file is designated by the inventors as "ds.cpp" in its uncompiled version. This file provides overall control to the entire OligoProbe DesignStation invention. It is divided into six sections. Section 0 defines and manipulates global variables. Section 1 controls general variable definition and initialization (including the arrays and memory blocks). It also reads and writes  
25 buffers for user input selections, and constructs multi buffers.

Section 2 sets up and initializes various "snippet" variables (see section below for a complete definition of the term snippet), converts base pair characters to a representation that is 96 base pairs long and to ASCII base pair strings, and performs other sequence file manipulation such as comparing snippets. This section also reads the sequence format file.  
30 reads base pairs, checks for and extracts sequence identification information (such as ORIGIN and LOCUS) and filters out sequences beginning with numbers.

Section 3 involves preparation file manipulation. This section performs the preprocessing on the PRP file discussed above. It also merges and sorts the snippet files.



-98-

creates a PRP file and sorts it, and outputs the sorted snippets. Next, this section streams through the PRP file.

5 Section 4 contains the essential code for H-Site Model processing (see Figures 41 through 43 for details, discussed below). Streams are set up, and then RIBI comparisons are performed for hybridizations (see file "ribi.cpp" for definitions of RIBI search techniques). Next, probes are generated, binding strength is converted to melting temperature, and hybridizations are calculated and stored (including hybridization strength). Lastly, other H-Site calculations are performed.

10 Section 5 is concerned with formatting and presenting diagnostic and user file (test.out, test1.out, and test2.out files) output. This section also handles the graphing functions (the MPSD diagram in particular). In addition, this section calculates the hairpin characteristics for the H-Site Model candidate probes.

15 The second H-Site Model file, designated as "ds.h" defines data variables and structures. Section 1 of this file concerns generic data structures (including memory blocks and arrays, and file inputs and outputs). Section 2 defines the variables and structures used with sequences, probes and hybridizations. Section 3 defines variables and structures concerned with protocols (i.e., function prototypes, graphing, etc.).

20 The third H-Site Model file, designated as "funcdoc.txt", contains very detailed documentation for this implementation of the H-Site Model program. Numerous variables and structures are also defined. The flow of the program is clearly shown in this file.

25 The fourth H-Site Model file, designated as "ribi.h" handles the sequence comparisons. The fifth and last H-Site Model file, designated as "ribi.cpp", performs internal B-Tree indexing. Definitions of Red-black Internal Binary Index (RIBI) searching are found in this file. Definitions are also included for the concepts keyed set, index, binary tree, internal binary index, paths, and red-black trees. Implementation notes are also included in this file.

#### d. Processing

30 Implementation of the H-Site Model in this invention is done in three stages. First, the invention creates the preparation (PRP) file, which contains all relevant information from the sequence database. This is the preprocessing stage discussed above. Next, the target is prepared by the program. Lastly, the invention calculates the MPSD data using the PRP file and target sequence to find probes.

i. Creation of the Preprocessed Preparation File

Figure 41. Step 1: The program first opens the sequence database for reading into memory 461. 462. Step 2: Next, as sequence base pairs are read in 462, "snippets" are saved to disk 463, along with loci information. A snippet is a fixed-length subsequence of a preparation sequence. The purpose of snippets is to allow the user to examine a small portion of a preparation sequence together with its surrounding base pairs. Snippets in the implementation of this invention are 96 base pairs long (except for snippets near the end or beginning of a sequence, which may have fewer base pairs). The "origin" of the snippet is in position 40. For snippets taken near the beginning of a sequence, some of the initial 40 bases are undefined. For snippets near the end of a sequence, some of the final 55 bases are undefined. Snippets are arranged in the preparation file (PRP) in sorted order (lexicographical order beginning at position 40). In this invention, the term "lexicographical order" means a preselected order, such as alphabetical, numeric or alphanumeric. In order to conserve space, snippets are only taken at every 4th position of the preparation sequence.

Step 3: The snippets are merge sorted 464 to be able to search quickly for sequences which pass the "screen", discussed below. Step 4: The merged file is prepended with identifiers for the sources of the snippets 465. This is done to identify the loci from which hybridizations arise.

ii. Target Preparation

Figure 42. Step 1: The target sequence file is opened 471 and read into memory 472. For each position in the target mRNA, the probe defined at that starting position is the shortest subsequence starting at that position whose hybridization strength is greater than the user specified melting temperature  $T_m$ . Typically, the probes are of length 18 to 50. Step 2: Four lists of "screens" are formed 473, 474, 475, each shifted by one base pair 475 to correspond to the fact that snippets are only taken at every four base pairs. A screen is a subsequence of the target mRNA of length equal to the screening threshold specified by the user. The screens are then indexed 476 and sorted in memory 477.

iii. Calculation of the MPSD Data

Figure 43. Step 3: This step is the heart of the process. Step 3a: The program streams through the following five items in sync, examining them in sequential order: the snippet file and the four lists of screens 481-484. Step 3b: Each snippet is compared to a screen 485. Step 3c: If the snippet does not match, whichever stream is behind is advanced 486 and Step 3b is repeated. If the snippet does match, Step 4 is performed.

-100-

Step 4: If a snippet and a matching screen were found in Step 3b 487, the hybridization strength of the binding between the sequence containing the snippet and all of the probes containing the screen is calculated (see Step 5). Double counting is avoided by doing this only for the first matched screen containing the probe. Each pair of bases is examined and assigned a numerical binding strength. An AT pair would be assigned a lower binding strength than a GC pair because AT pairs have a lower melting temperature  $T_m$ . The process is explained more fully below at Step 5b.

Step 5: The hybridization strengths between sequence and all the probes containing it are calculated using a dynamic programming process. The process is as follows: Step 5a: Begin at the position of the first probe containing the given screen but not containing any other screens which start at an earlier position and also match the sequence. This is done to avoid double counting. Two running totals are maintained: a) boundStrength, which represents the hybridization strength contribution which would result if the sequence and probe were to match exactly for all base pairs to the right of the current position, and b) unboundStrength, which represents the strength of the maximally binding region. Step 5b: At each new base pair, the variable boundStrength is incremented by 71 if the sequence and probe match and the matched base pair is GC 489, incremented by 30 if the matched base pair is AT 490 (i.e., this number is about 42.25% of the first number 71), and decremented by 74.5 if there is not a match 488 (i.e., this number is about 5% larger than the first number 71). Step 5c: If the current boundStrength exceeds the current unboundStrength 491 (which was originally initialized to zero), a new binding region has been found, and unboundStrength is set equal to boundStrength 492. Step 5d: If the current boundStrength is negative, boundStrength is reset to zero 493. Step 5e: If the current position is at the end of a probe, the results (the hybridization strengths) are tallied for that probe. Step 5f: If the current position is at the end of the last probe containing the screen, the process stops.

Step 6: A tally is kept of the number and melting temperature of the matches for each candidate probe, and the location of the best 20 candidates, using a priority queue (reverse order by hybridization strength number) 494. Step 7: A numerical "score" is kept for each preparation sequence by tallying the quantity  $\exp(-T_m)$  (which can be expressed as  $\Sigma e^{-T_m}$ ) for each match 495, where  $T_m$  is the melting temperature for the "perfect" match, the probe itself. In other words, the probe hybridizes "perfectly" to its target.

Step 8: Hairpins are calculated by first calculating the complementary probe. In other words, the order of the bases in the candidate probe are reversed (CTATAG to

-101-

GATATC), and complementary base pairs are substituted (A for T, T for A, G for C, and C for G, changing GATATC to CTATAG in the above example). Next, the variable representing the maximum hairpin length for a candidate probe is initialized to zero, as is the variable representing a hairpin's distance. For each offset, the original candidate probe and the complementary probe just created are then aligned with each other and compared. The longest match is then found. If any two matches have the same length, the one with the longest hairpin distance (i.e., the number of base pairs separating the match) is then saved.

Step 9: The preparation sequences are then sorted 496 and displayed in rank order, from best to worst 497. Step 10: The resulting MPSD, which includes all candidate probes, is then displayed on the screen. Step 11: The best 20 matches are also printed or displayed in rank order, as the user requests 497.

e. Outputs

The outputs of the H-Site Model are fully described in Section 1(d)(iv), above, and illustrated in Figures 23 through 25. Samples of the two output files created by the H-Site Model are shown in Figures 44A and 44B.

4. Description of the Mitsuhashi Probe Selection Diagram Processing

Once the Mitsuhashi Probe Selection Diagram (MPSD) data has been calculated by the H-Site Model program (see stage three and Figure 43, discussed above), it is necessary to convert this data to pixel format and plot a graph. An overview of this process is shown in Figure 45. First, the program calculates the output (x,y) ranges 500. Next, these are converted to a logarithmic scale 501. The values are then interpolated 502, and a bitmap is created 503. Lastly, the bitmap is displayed on the screen 504 in MPSD format (discussed above in section 1(e)(i)). A sample MPSD is shown in Figure 23.

5. Description of the MatchInfo Window Processing

The ProbeInfo and MatchInfo windows are discussed in great detail in Section 1(e)(ii), and a sample of these windows is shown in Figure 24. An overview of the processing involved in creating the MatchInfo portion of the window is given in the flow chart in Figure 46. First, as the user moves the MPSD cursor 570 (seen as a vertical line bisecting the MPSD window), the program updates the position of the candidate probe shown under that cursor position 521. Next, based upon the candidate probe's position, the program updates the sequence 522 and hairpin information 523 for that probe. This updated information is then displayed in an updated match list 524, shown in the MatchInfo window.

**Effect of the Invention**

The invented method offers rapid characterization of mRNA in approximately 5 hours without the need for mRNA purification. Because conventional methods of mRNA quantification, for example Northern blots, require over 70 hours to complete, the method  
5 of the present invention significantly reduces the length of time needed for mRNA analysis. Furthermore, because various types of first nucleotide probes can be immobilized in each well of microtiter plate and identified using a single kind of labeled second nucleotide probe, the present invention advantageously can provide the simultaneous  
10 quantification of multiple varieties of mRNA.

As shown in the examples above, the methods of the present invention provide high reliability and reproducibility. Therefore, these methods can be utilized for quantification and analysis of many mRNAs in addition to those exemplified here. Thus, this invention can be used as a rapid method of analyzing the pathogenesis of various diseases, as well as assisting in their diagnosis. Many such methods will suggest themselves to those with  
15 ordinary skill in the art upon a review of the present specification. Accordingly, the scope of the present invention is to be determined with reference to the appended claims, and is not to be limited to those methods specifically exemplified herein.

-103-

SEQUENCE LISTING

## (1) GENERAL INFORMATION:

5

(i) APPLICANT: Akitaya, Tatsuo  
Mitsubishi, Masato  
Cooper, Allan

10

(ii) TITLE OF INVENTION: METHOD AND REAGENT FOR MEASURING  
MESSENGER RNA

(iii) NUMBER OF SEQUENCES: 711

15

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: KNOBBE, MARTENS, OLSON, AND BEAR  
(B) STREET: 620 NEWPORT CENTER DRIVE SIXTEENTH FLOOR  
(C) CITY: NEWPORT BEACH  
(D) STATE: CA  
(E) COUNTRY: USA  
(F) ZIP: 92660

20

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk  
(B) COMPUTER: IBM PC compatible  
(C) OPERATING SYSTEM: PC-DOS/MS-DOS  
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

25

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:  
(B) FILING DATE:  
(C) CLASSIFICATION:

30

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Altman, Daniel E  
(B) REGISTRATION NUMBER: 34,115  
(C) REFERENCE/DOCKET NUMBER: HITACHI.006CP2

35

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: 714-760-0404  
(B) TELEFAX: 714-760-9502

40

## (2) INFORMATION FOR SEQ ID NO:1:

45

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 bases  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

50

-1(1--

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

5 (iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Gs mRNA NUCLEOTIDE PROBE

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

TTCATCCTCC CACAGAGCCT TG

22

15

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 bases

(B) TYPE: nucleic acid

20 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

25

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

30

(A) ORGANISM: Gi-2 mRNA NUCLEOTIDE PROBE

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

35

ATGGTCAGCC CAGAGCCTCC GG

22

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 bases

(B) TYPE: nucleic acid

40 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

45

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

50

-105-

(vi) ORIGINAL SOURCE:

(A) ORGANISM: G PROTEIN SENSE PCR PRIMER

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

AGCACCATTG TGAAGCAGAT GA

22

(2) INFORMATION FOR SEQ ID NO:4:

10

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 bases

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

15

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

20

(iv) ANTI-SENSE: YES

(vi) ORIGINAL SOURCE:

(A) ORGANISM: G PROTEIN ANTISENSE PCR PRIMER

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

CTCTGGCCTC CCACATCAAA CA

22

30

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 bases

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

35

(ii) MOLECULE TYPE: cDNA to mRNA

40

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: SUBSTANCE P RECEPTOR mRNA PROBE

45

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:



-106-

GGACTTATGA GAAAGCGTAC CA

22

## (2) INFORMATION FOR SEQ ID NO:6:

## 5 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 bases

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

15 (iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: HUMAN B2 RECEPTOR mRNA PROBE

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

ATGCTGGCCG TGACGCACA GCA

23

25

WE CLAIM:

1. A highly sensitive, quantitative and rapid method for detecting and quantifying mRNA in a sample without the need to purify mRNA from cells, said method comprising:

- 5 (a) identifying a polynucleotide sequence that is unique to said mRNA;
- (b) immobilizing a first polynucleotide to an insoluble support, said first polynucleotide having a first sequence complementary to said sequence unique to said mRNA;
- 10 (c) incubating said sample with said insoluble support under conditions wherein said unique sequence will hybridize with said first polynucleotide, thereby immobilizing mRNA present in said sample to said insoluble support;
- (d) washing non-immobilized components of said sample from said insoluble support;
- 15 (e) labeling mRNA on said support in a manner that label is incorporated onto said support related to the amount of mRNA on said support; and
- (f) measuring the amount of label immobilized on said support.

2. The method of Claim 1, wherein step (e) comprises:

- 20 i. incubating a labeled probe with said insoluble support under conditions wherein said labeled probe will hybridize with mRNA on said support, said labeled probe bearing a label and being complementary to a part of said mRNA other than a part of said mRNA complementary to said first probe, thereby immobilizing to said insoluble support the label on said labeled probe which has hybridized with said mRNA; and
- 25 ii. washing any non-immobilized labeled probe from said insoluble support.

3. The method of Claim 2, wherein said labeled probe comprises poly-d(T).

4. The method of Claim 1, wherein step (e) comprises labeling oligonucleotides on said support with a nucleic acid stain.

30 5. The method of Claim 4, wherein said nucleic acid stain is selected from the group consisting of ethidium bromide, yoyo-1 and toyo-1.

6. The method of Claim 2, wherein said label comprises a radionuclide.

7. The method of Claim 2, wherein said label comprises biotin.

8. The method of Claim 7, wherein step (f) comprises:

-108-

incubation with enzyme-conjugated streptavidin; and  
measurement of the amount of enzyme bound to said insoluble support.

9. The method of Claim 1, wherein said mRNA codes for a  $\beta$  receptor, and wherein said first polynucleotide comprises the sequence of SEQ ID NO:6.

5 10. The method of Claim 1, wherein said mRNA codes for the  $\alpha$  subunit of Gs protein, and wherein said first polynucleotide comprises a sequence complementary to a sequence selected from the group consisting of SEQ ID NO:1, 145-146, 209-212, 268-287, 705 and 708.

10 11. The method of Claim 1, wherein said mRNA codes for Gi-2 protein, and wherein said first polynucleotide comprises a sequence complementary to a sequence selected from the group consisting of SEQ ID NO:703, 806, 142, 143, 187-197 and 238-253.

12. The method of Claim 1, wherein said mRNA codes for substance P receptor, and wherein said first polynucleotide comprises the sequence of SEQ ID NO:5.

15 13. The method of Claim 1, wherein said mRNA codes for Gi-3 protein, and wherein said first polynucleotide comprises a sequence complementary to a sequence selected from the group consisting of SEQ ID NO:704, 707, 144, 198-208 and 254-267.

14. The method of Claim 1, wherein said mRNA codes for Gi-1 protein, and wherein said first polynucleotide comprises a sequence complementary to a sequence selected from the group consisting of SEQ ID NO:702, 805, 141, and 224-237.

20 15. The method of Claim 1, wherein said mRNA codes for G protein, and wherein said first polynucleotide comprises a sequence complementary to a sequence selected from the group consisting of SEQ ID NO:700, 701, 139, 140, 148-159, 162-172, 176-186 and 3-4.

25 16. The method of Claim 1, wherein said first sequence hybridizes to said unique sequence of said mRNA and not to other polynucleotides present in said sample.

17. The method of Claim 1, wherein said sample comprises a cell lysate.

18. The method of Claim 2, wherein said label can be measured by light emitted therefrom, and wherein step (f) comprises measuring the amount of light emitted by said label.

30 19. The method of Claim 18, wherein said label comprises biotin and wherein step (f) additionally comprises adding streptavidin bound to alkaline phosphatase to said insoluble support.

20. The method of Claim 18, wherein the step of measuring the amount of light comprises:

recording the amount of light on film; and  
measuring the exposure of this film using a densitometer.

21. The method of Claim 18, additionally comprising adding ATTOPHOS and measuring fluorescence emitted using a fluorimeter.

5 22. The method of Claim 21, wherein the step of measuring the amount of light comprises use of a fluorimeter.

23. The method of Claim 1, wherein step (a) comprises use of a computer program.

10 24. The method of Claim 23, wherein step (a) comprises use of an H-site model.

25. The method of Claim 24, wherein identification of the polynucleotide sequence using said H-site model comprises the steps of:

specifying a minimum melting temperature for the first polynucleotide and said polynucleotide sequence;

15 specifying a nucleation threshold that places a minimum value on the number of base pairs at any nucleation site;

determining the melting temperatures ( $T_m$ ) of the first polynucleotide and said polynucleotide sequence at every possible hybridization point; and

selecting the nucleotide sequence having the highest  $T_m$  value.

20 26. The method of Claim 25 wherein said melting temperature is determined by the formula:  $T_m = 81.5 - 16.6(\log[Na]) - 0.63\%(\text{formamide}) + 0.41(\%(G+C)) - 600/N$ , wherein  $\log[Na]$  is the log function of the sodium concentration,  $0.063\%(\text{formamide})$  is the concentration of formamide,  $\%(G+C)$  is the percentage of matched GC base pairs, and N is the probe length.

25 27. The method of Claim 2, wherein the labeled probe is identified using an H-site model comprising the steps of:

specifying a minimum melting temperature for the labeled probe and the part of said mRNA complementary to a part of said mRNA other than a part of said mRNA complementary to said first probe;

30 specifying a nucleation threshold that places a minimum value on the number of base pairs at any nucleation site;

determining the melting temperatures ( $T_m$ ) of the first nucleotide probe and said part of said mRNA complementary to a part of said mRNA other than

-110-

a part of said mRNA complementary to said first probe at every possible hybridization point; and

selecting the nucleotide probe of the proper length having the lowest  $T_m$  value.

5           28. The method of Claim 27 wherein said melting temperature is determined by the formula:  $T_m = 81.5 - 16.6(\log[\text{Na}]) - 0.63(\%(\text{formamide})) + 0.41(\%(G + C)) - 600/N$ , wherein  $\log[\text{Na}]$  is the log function of the sodium concentration,  $0.063\%(\text{formamide})$  is the concentration of formamide,  $\%(G + C)$  is the percentage of matched GC base pairs, and N is the probe length.

10           29. A method of inhibiting RNase activity in a sample containing RNA in a buffer system containing SDS and EDTA, comprising adding vanadyl ribonucleoside complex to said sample.

30. The method of Claim 29, wherein the sample contains at least a five-fold molar excess of EDTA over RNA in the sample.

15           31. The method of Claim 29, wherein the vanadyl ribonucleoside complex is added in combination with proteinase K.

32. The method of either Claim 29 or 31, wherein the vanadyl ribonucleoside complex is added in combination with RNasin.

20           33. A polynucleotide-immobilized support that is useful in detecting and quantifying the amount of  $\beta$  receptor mRNA in a sample, comprising:

an insoluble support:

a first polynucleotide bound to said support, said first polynucleotide comprising a sequence that is complementary to a sequence unique to  $\beta$  receptor mRNA.

25           34. A polynucleotide-immobilized support as in Claim 33, wherein said sequence unique to  $\beta$  receptor mRNA comprises the sequence of SEQ ID NO:6.

35. A polynucleotide-immobilized support that is useful in detecting and quantifying the amount of jun oncogene mRNA in a sample, comprising:

an insoluble support:

30           a first polynucleotide bound to said support, said first polynucleotide comprising a sequence that is complementary to a sequence unique to jun oncogene mRNA.

36. A polynucleotide-immobilized support as in Claim 35, wherein said sequence unique to jun oncogene mRNA comprises a sequence complementary to a

sequence selected from the group consisting of SEQ ID NO:11-35, 39-53, 62-138, 310-342, 600-606, 614-620, and 628-634.

37. A polynucleotide-immobilized support that is useful in detecting and quantifying the amount of G protein mRNA in a sample, comprising:

5 an insoluble support;

a first polynucleotide bound to said support, said first polynucleotide comprising a sequence that is complementary to a sequence unique to G protein mRNA.

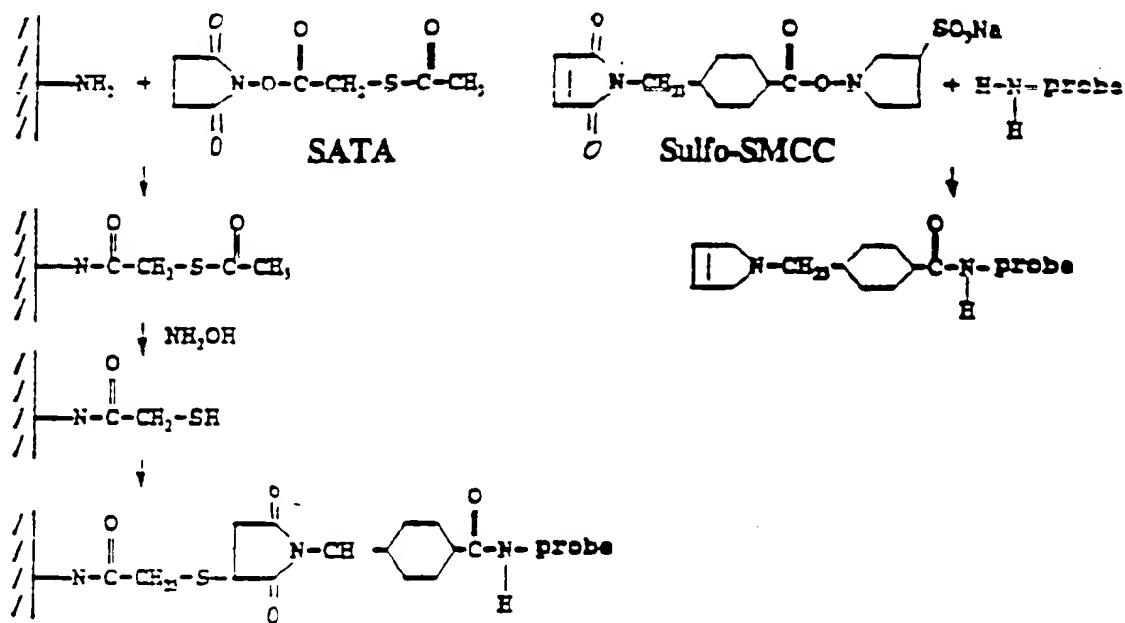
38. A polynucleotide-immobilized support as in Claim 37, wherein said  
10 sequence unique to G protein mRNA comprises a sequence selected from the group consisting of SEQ ID NO:1, 2, 700-709, 805, 806, 139-159, 162-172 and 176-309.

39. A polynucleotide-immobilized support that is useful in detecting and quantifying the amount of Substance P receptor mRNA in a sample, comprising:

15 an insoluble support;

a first polynucleotide bound to said support, said first polynucleotide comprising a sequence that is complementary to a sequence unique to Substance P receptor mRNA.

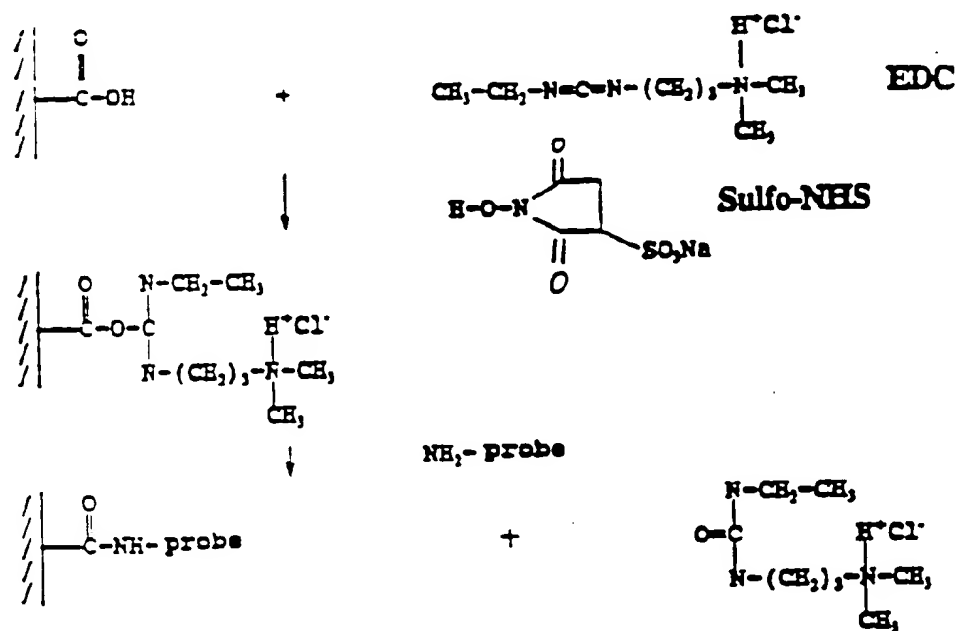
40. A polynucleotide-immobilized support as in Claim 39, wherein said  
20 sequence unique to Substance P receptor mRNA comprises the sequence of SEQ ID NO:5.



SATA: Succinimidyl-S-acetylthioacetate  
 Sulfo-SMCC: Sulfo-succinimidyl-4-(N-maleimidomethyl)  
 cyclohexane-1-carboxylate

Fig.1

2 / 59



EDC: 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride  
 Sulfo-NHS: N-hydroxysulfosuccinimide

Fig.2



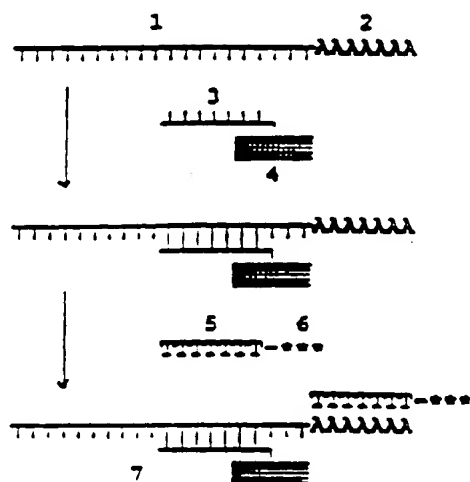


Fig.3

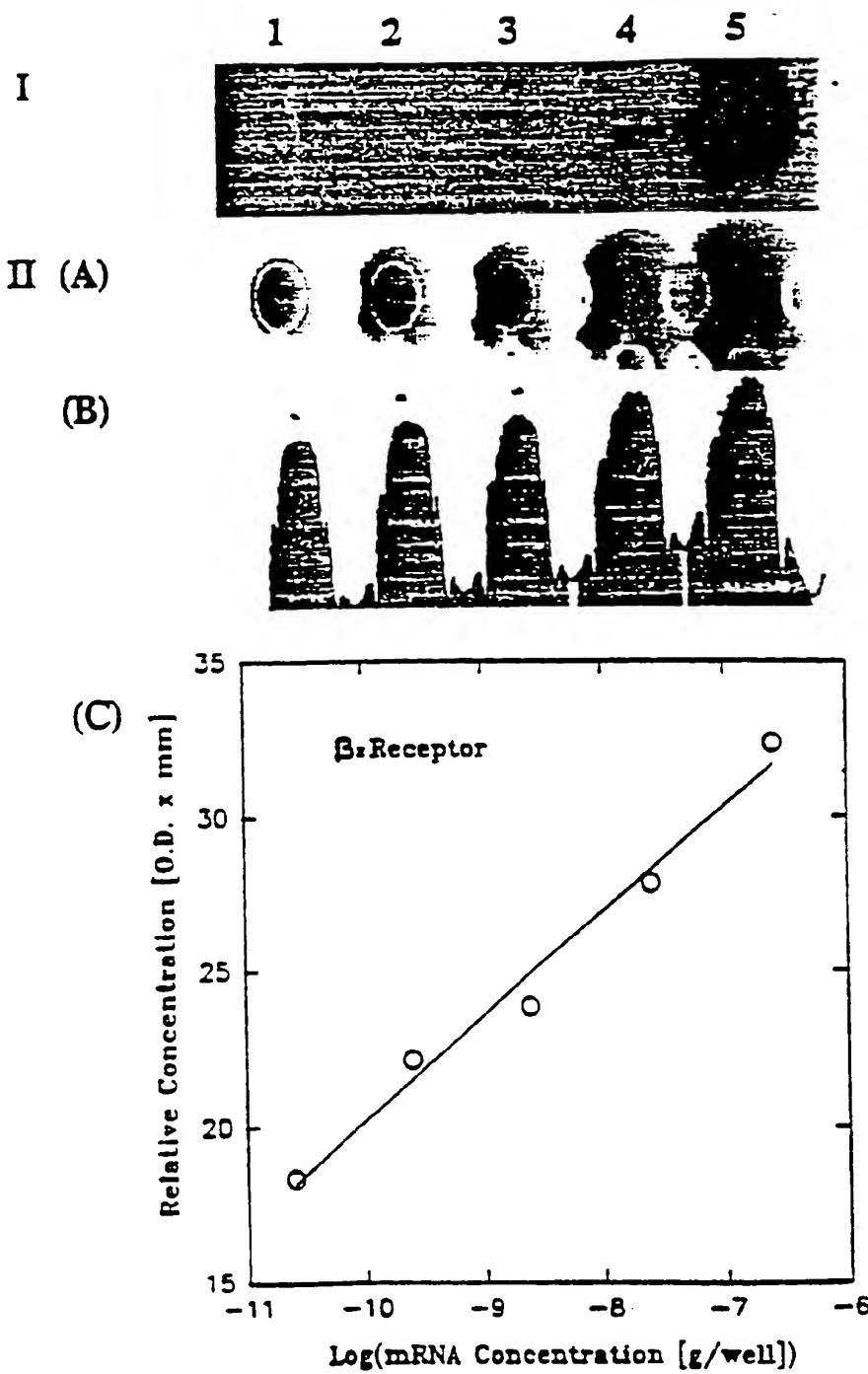


Fig. 4.

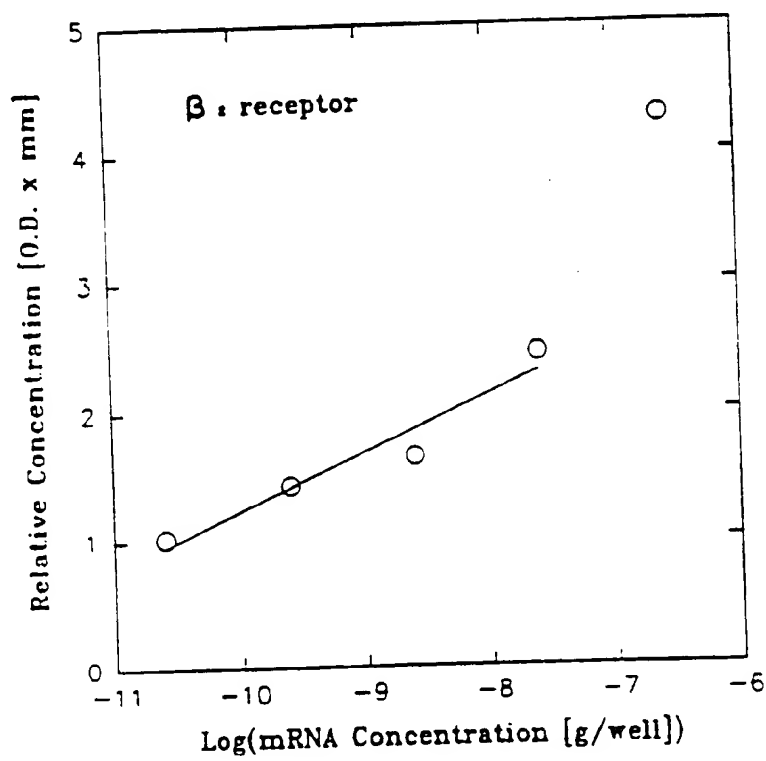


Fig.5

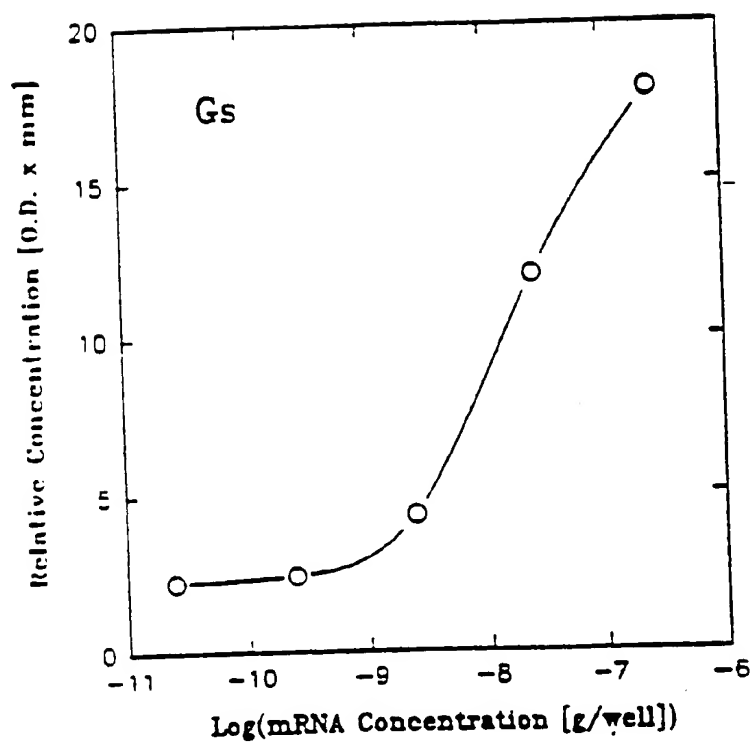


Fig.6

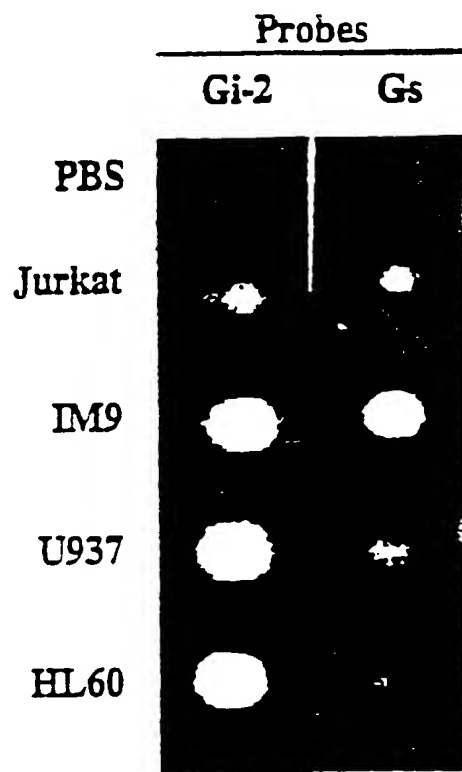


Fig.7

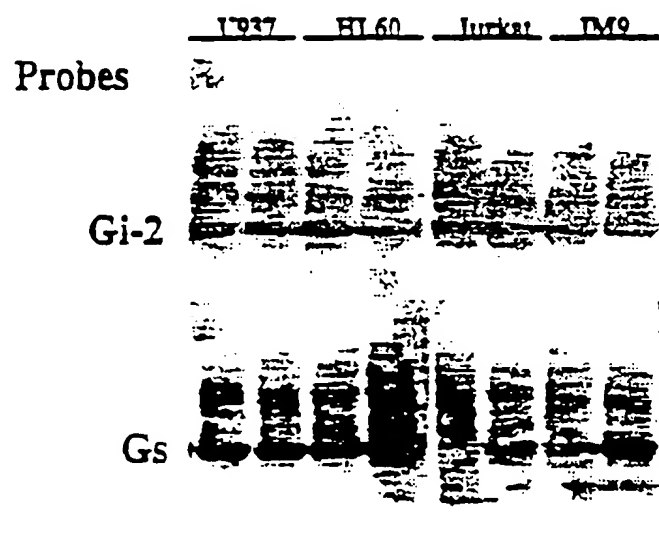


Fig.8

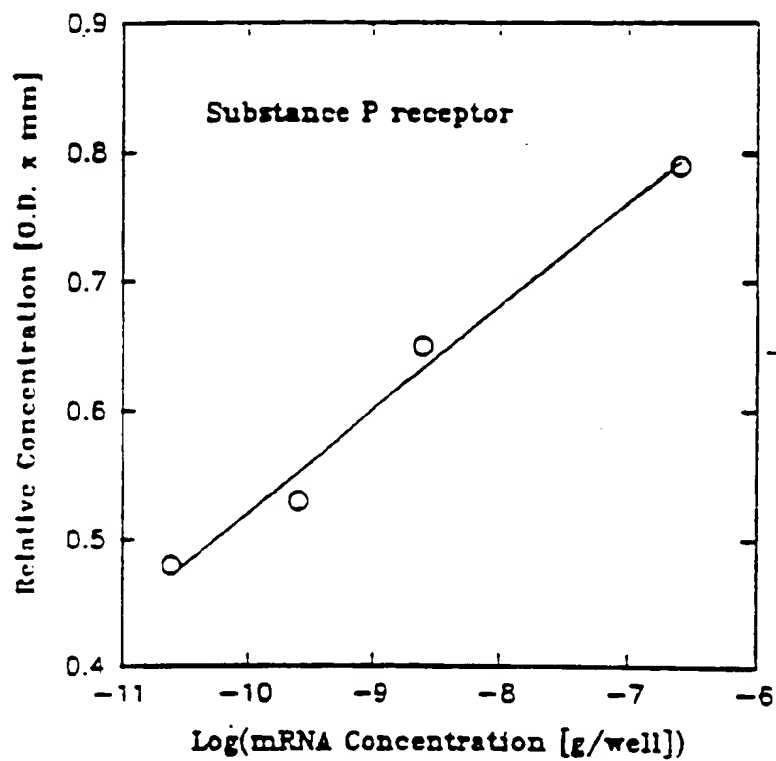


Fig.9

# Effect of various RNase inhibitors on mRNA preparation

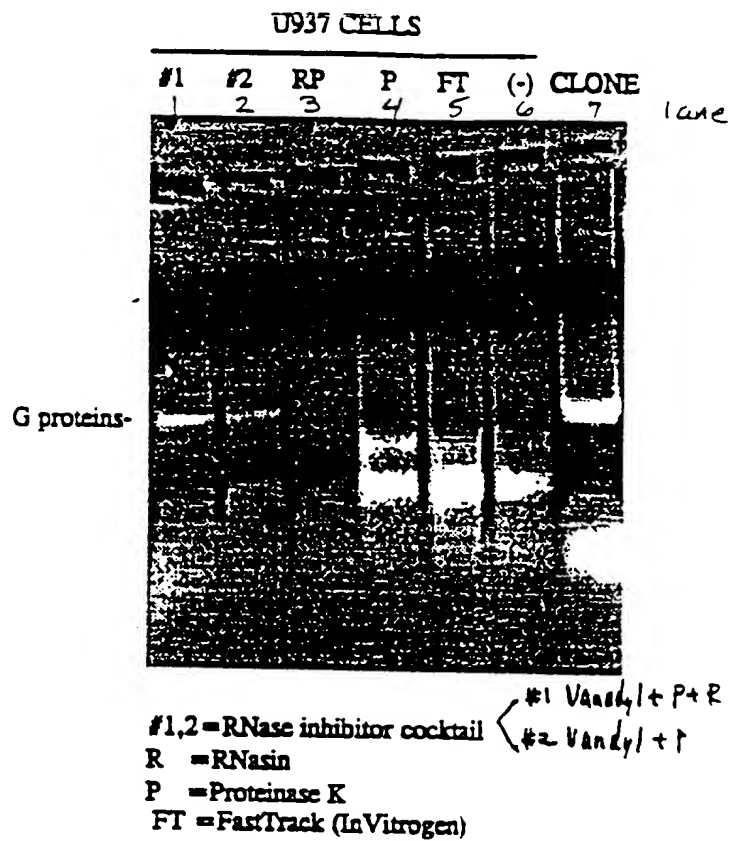


Fig. 10



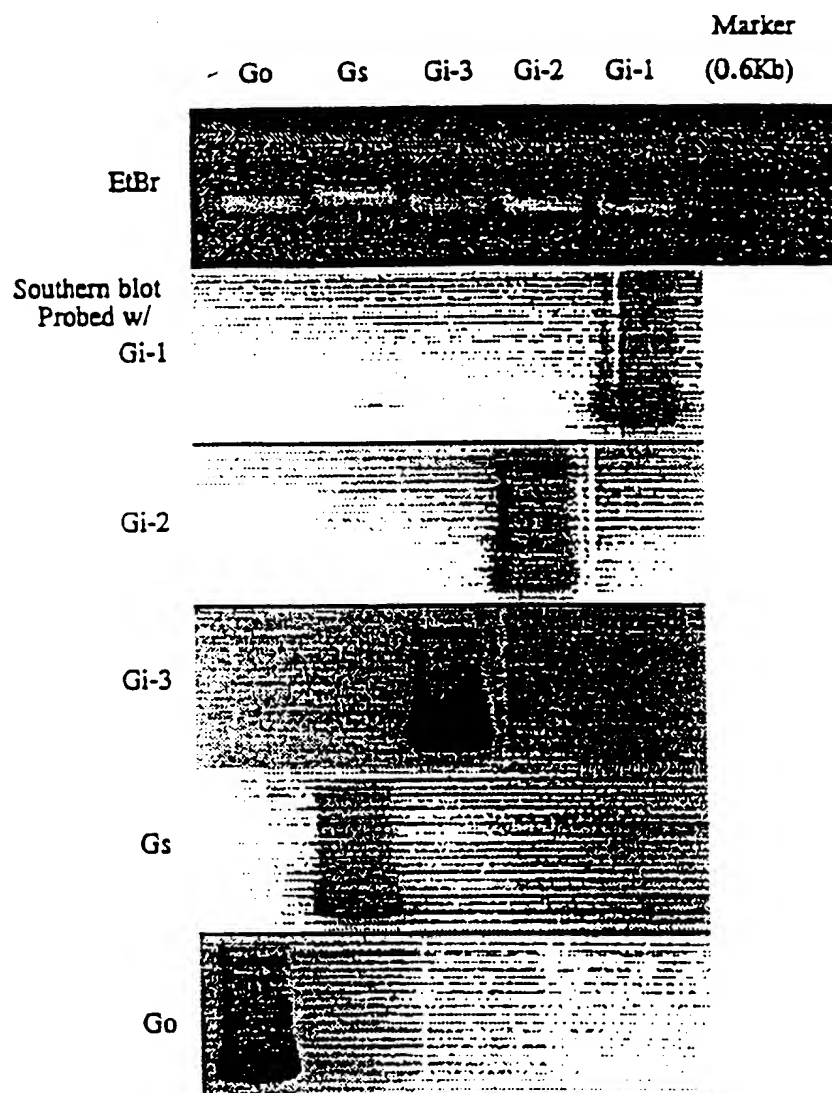


Fig. 11

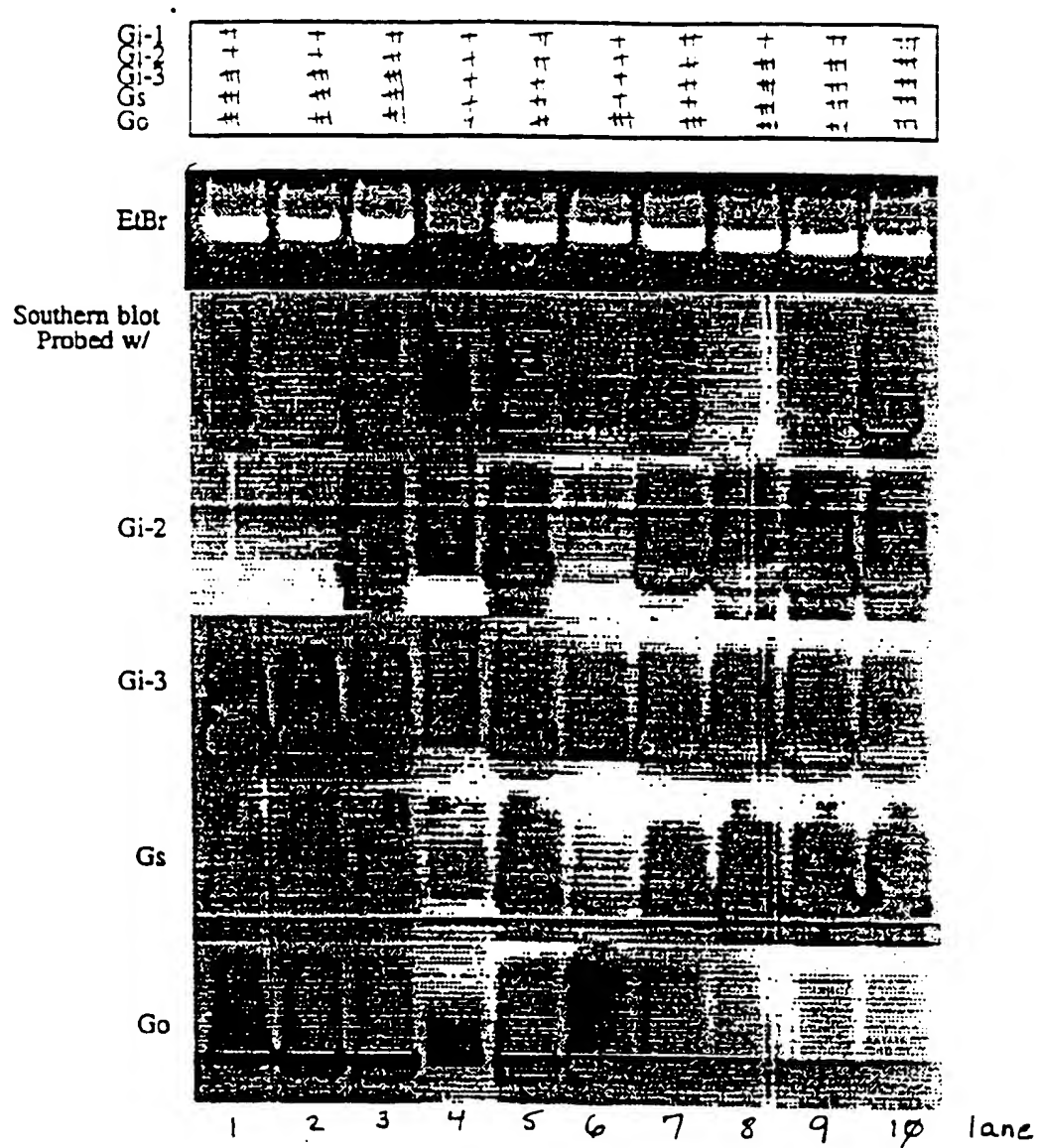


Fig. 12

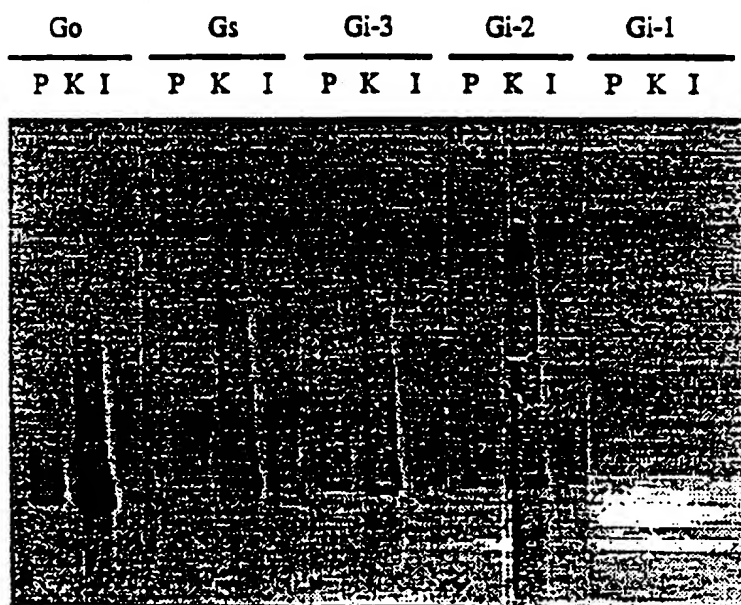


Fig. 13

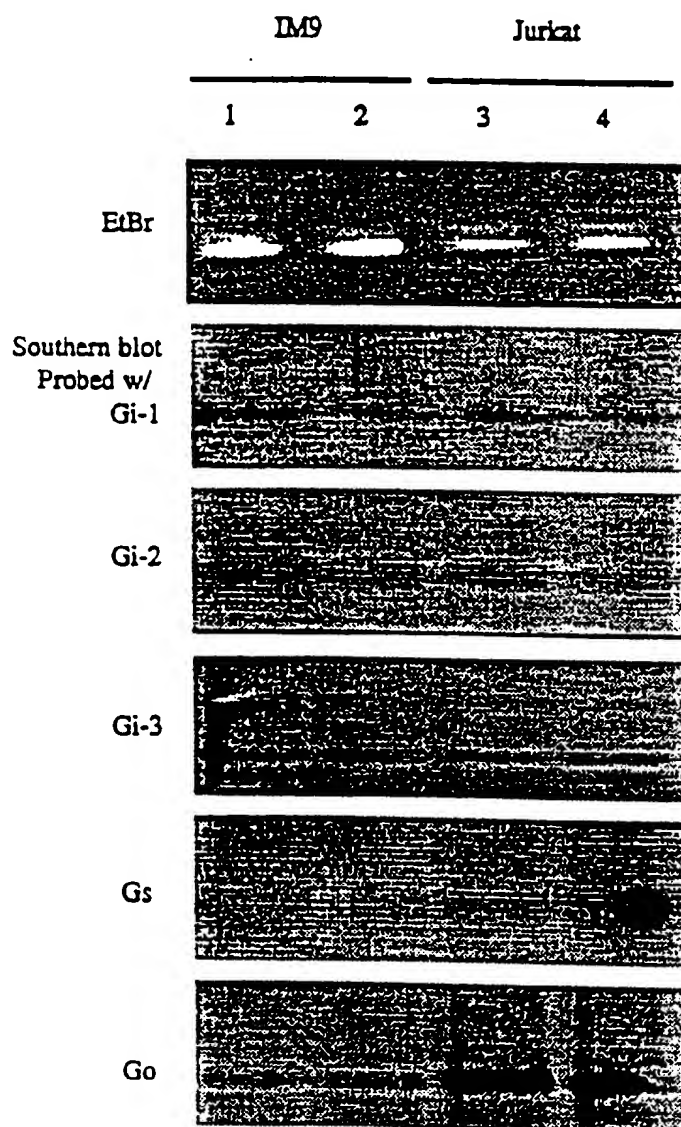


Fig. 14

15/59

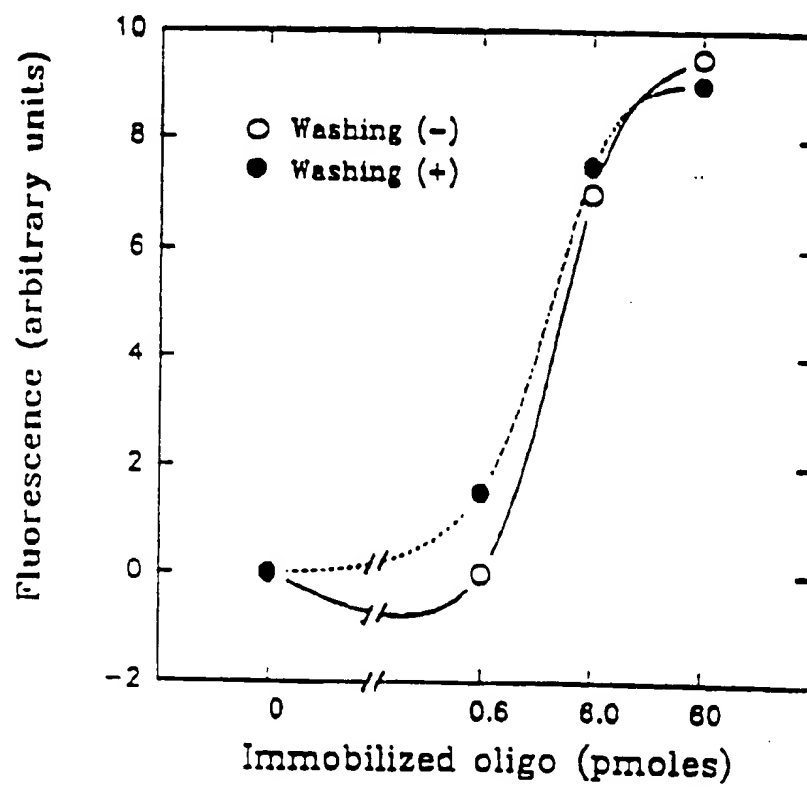


Fig. 15

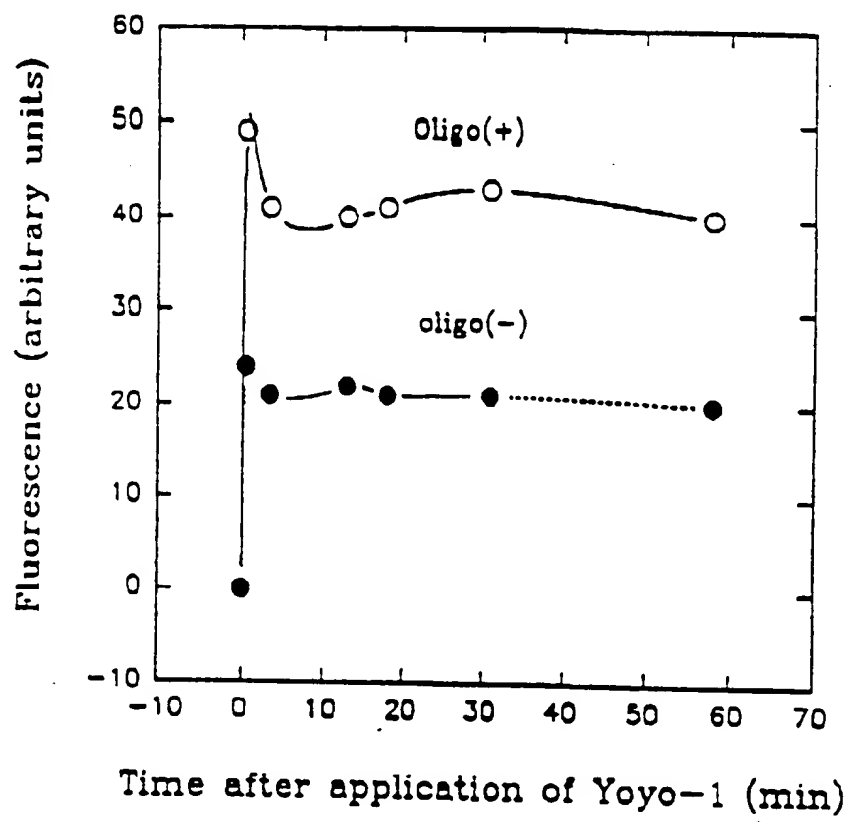


Fig. 16

17/59

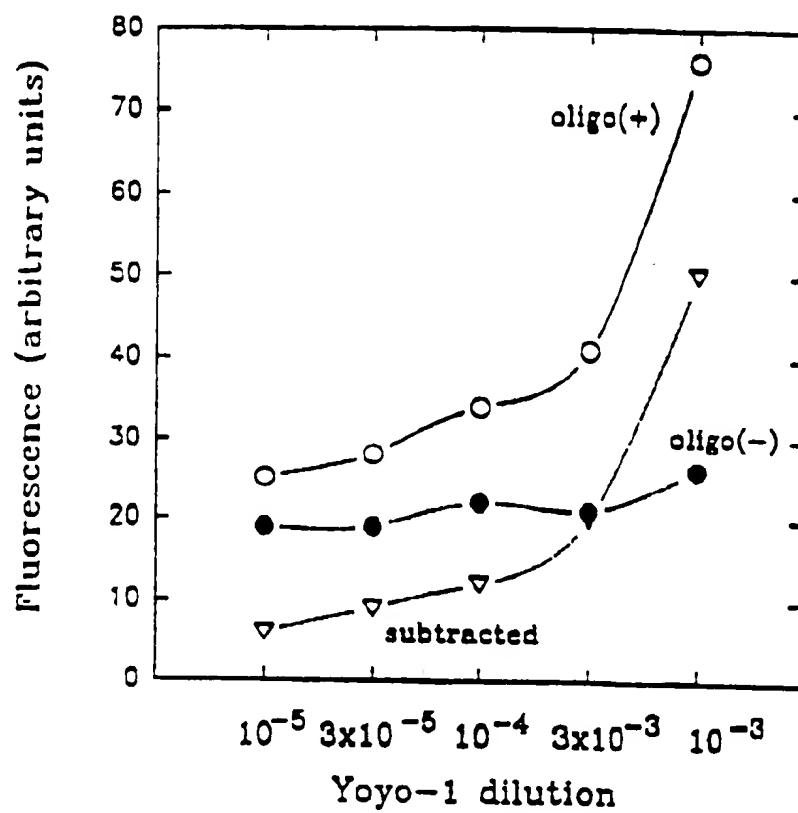


Fig. 17

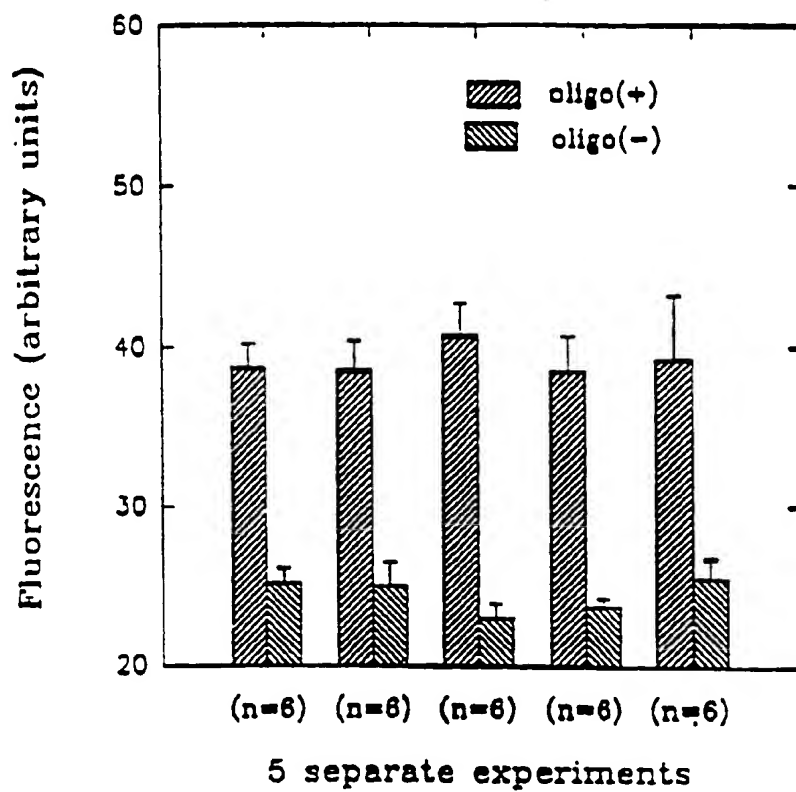


Fig. 18



19/59

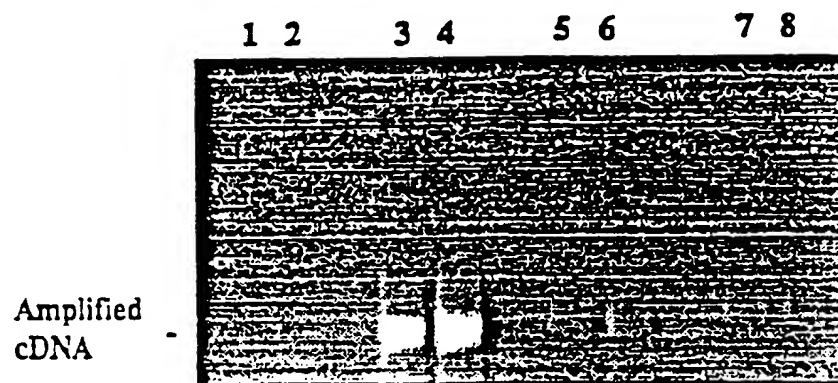


Fig. 19

FIG. 20

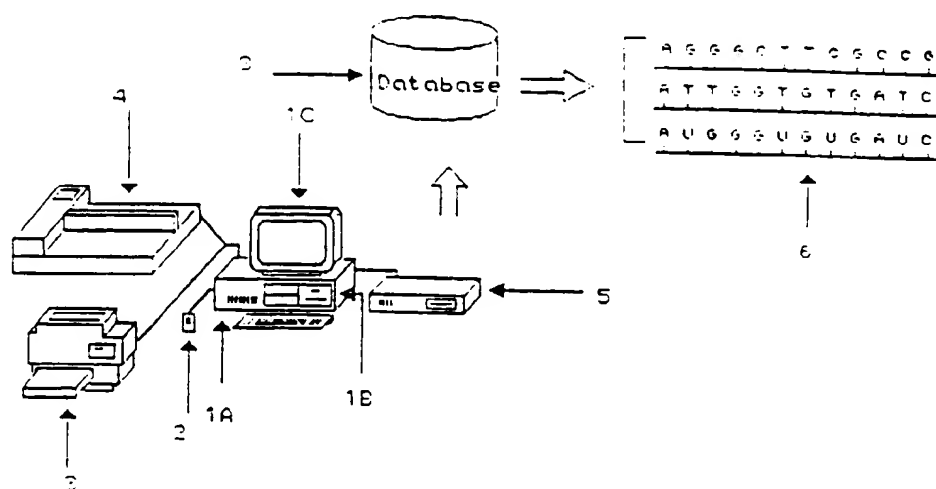
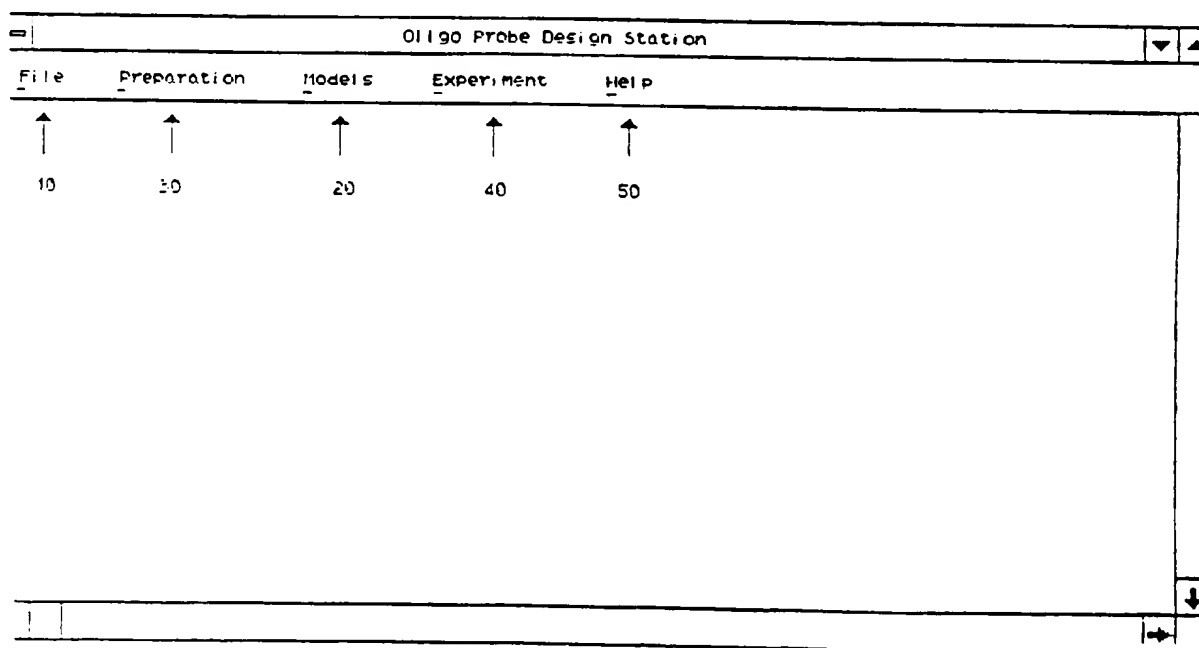


FIG. 21A



21/59

FIG. 21B

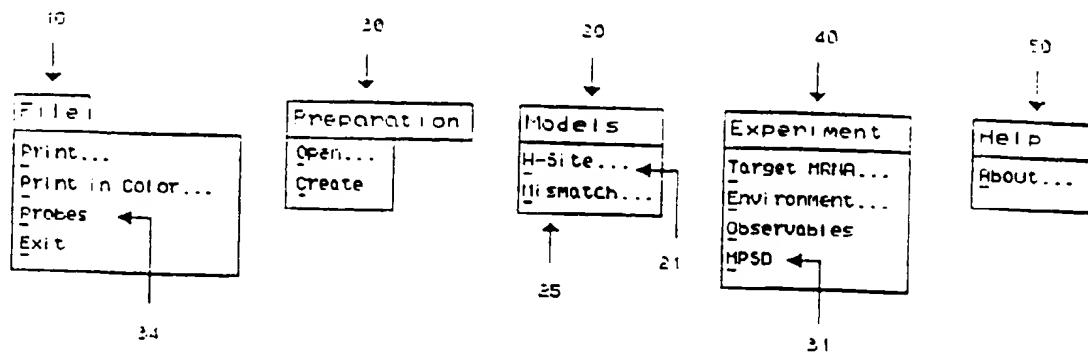


FIG. 21C

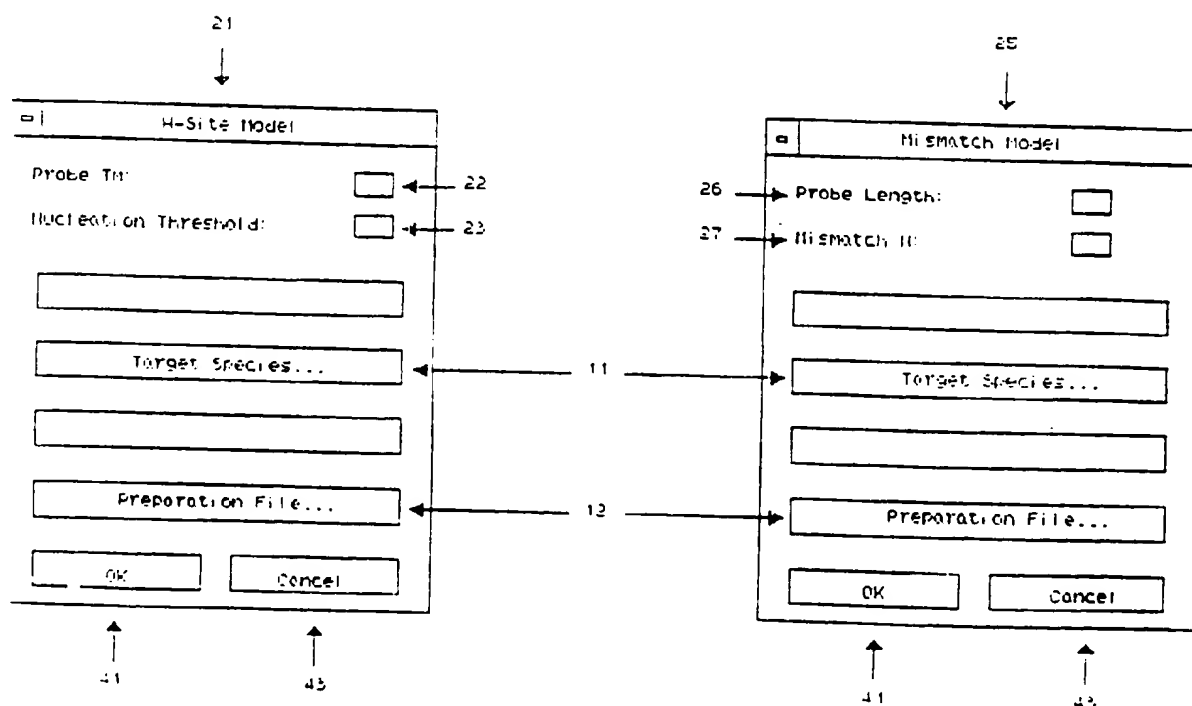


FIG. 22

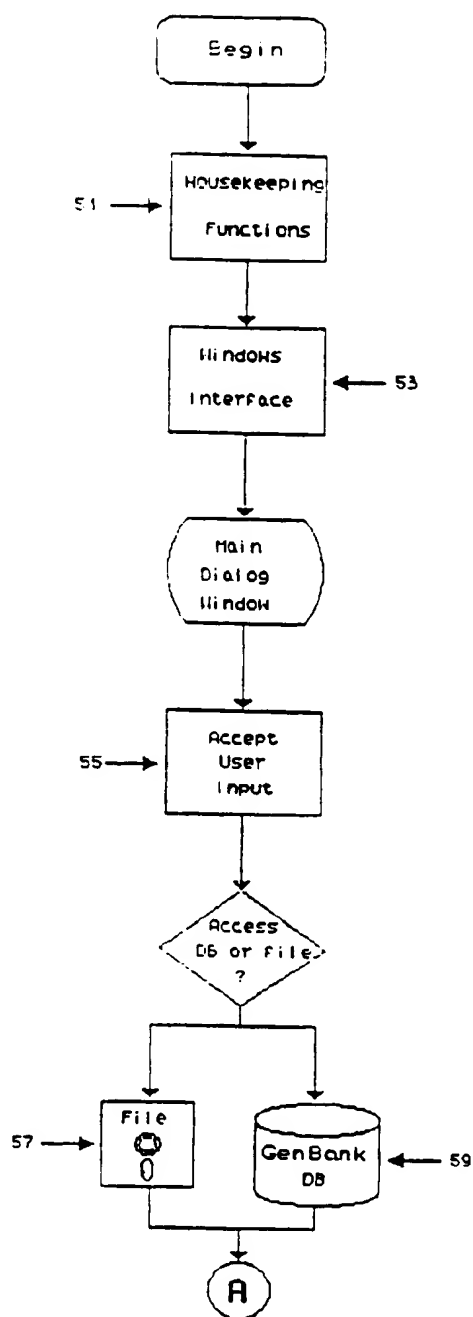


FIG. 22 CONT.

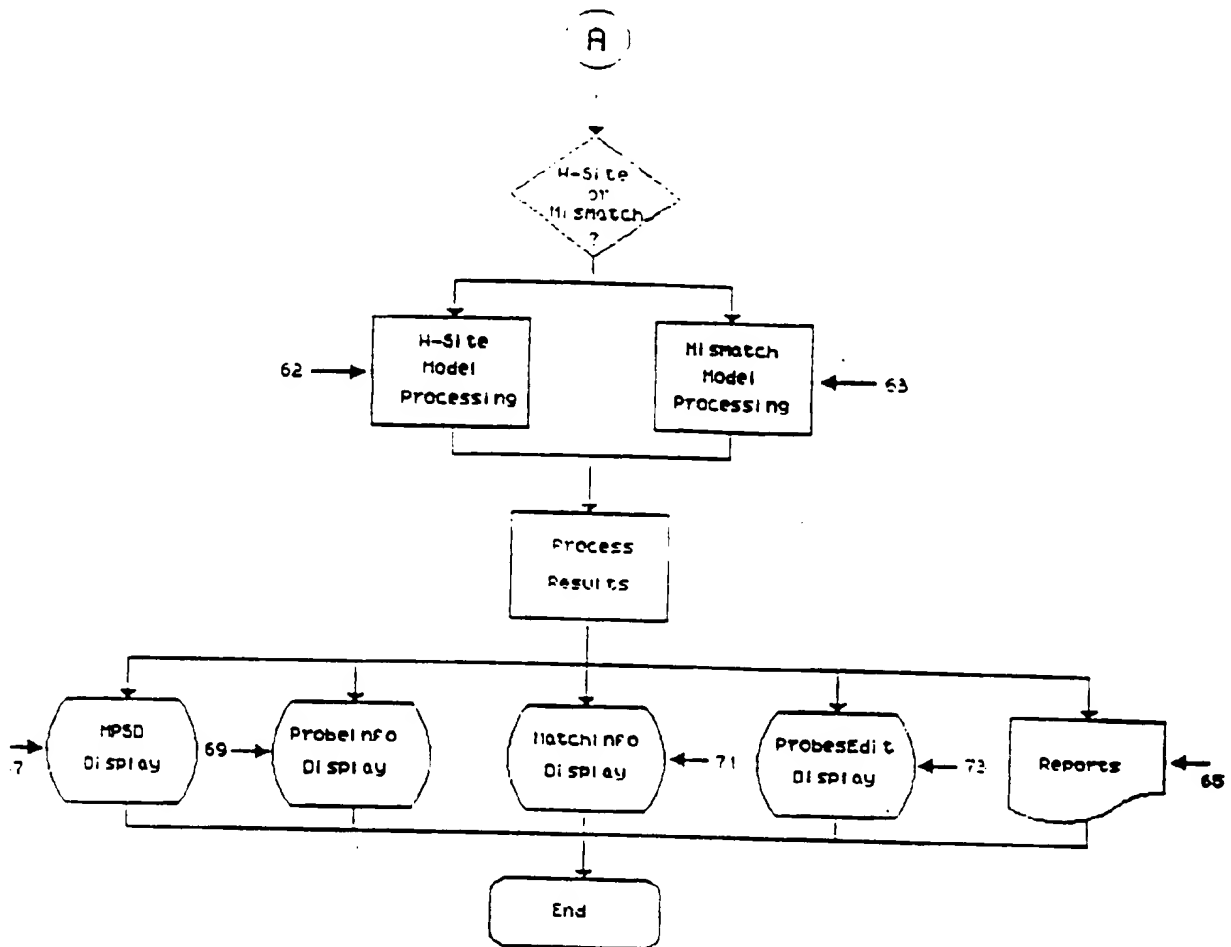


FIG. 23

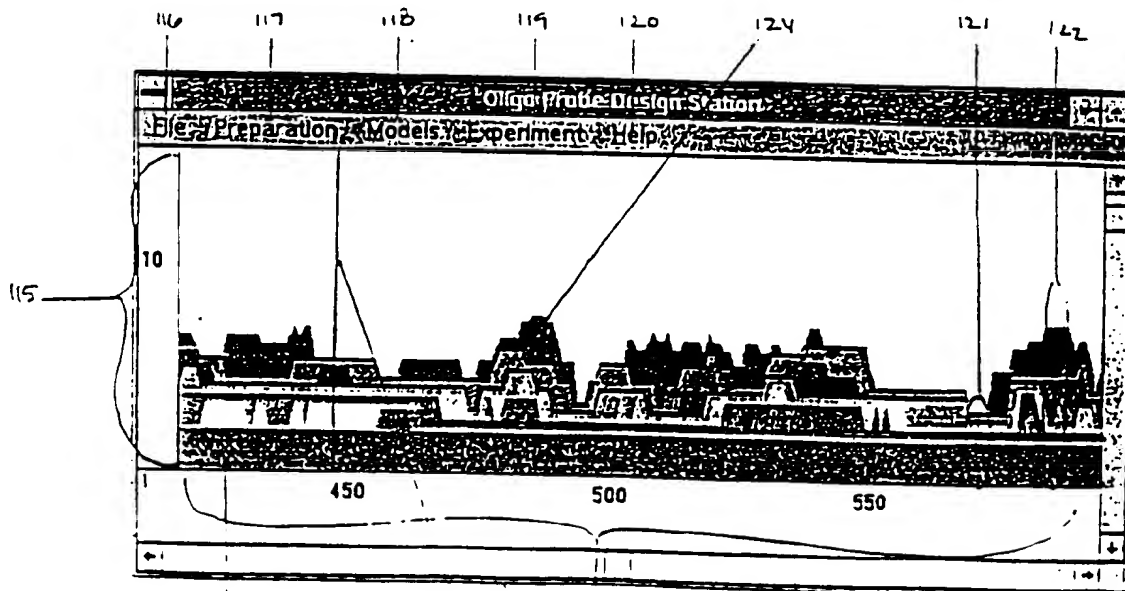


FIG. 24

ProbeInfo				
398	PROBE: F:\MILAN\HUMBUNX.CDS			
HYBRIDIZATIONS: F:\MILAN\JUNMIX.PRP				
Locus	Pos	Tm	Length	21
humbjunx	398	61.0	999cttcgcccgcggtttg	4
musbjunx	398	61.0	-----g-----	1
humbjunx	223	61.0	-----g-----	
humbjunx	215	61.0	-----g-----	
humbjunx	401	61.0	ccctgc-----gccc	
musbjunx	401	61.0	-----g-----	
humbjunx	100	61.0	g-----gc-----	
musbjunx	262	61.0	ct-----gc-----	
humbjunx	602	61.0	cc-----ggtct-----	
humbjunx	443	61.0	cc-----gc-----	
musbjunx	398	61.0	-----g-----	
musbjunx	398	61.0	-----g-----	
humbjunx	398	61.0	-----g-----	

FIG. 25

Probes Selected - JUNMIX.prb	
File	
PROBE: C:\HITACHI\JUNMIX.PRP	
HYBRIDIZATION: C:\HITACHI\HUMBUNX	
Length = 374 Hairpin = 3 5	
Locus	Pos Tm
humbjunx	374 61.47
musbjunx	365 61.47
humbjunx	41 34.82 t-----g-g
humbjunx	102 31.12 a-----gtgg
humbjunx	602 31.12 c-----c-gg
PROBE: C:\HITACHI\JUNMIX.PRP	
HYBRIDIZATION: C:\HITACHI\HUMBUNX	
Length = 467 Hairpin = 2 13	
Locus	Pos Tm
humbjunx	467 61.7
musbjunx	458 51.6
humbjunx	32 29.35 tgaagcgg
humbjunx	32 29.35 tgaagcgg

FIG. 26A

PROBE: C:\HITACHI\JUNMIX.PRP  
 HYBRIDIZATION: C:\HITACHI\HUMBJUNX.CDS  
 Length = 374 Hairpin = 3 5  
 Locus Pos Tm  
 humbjunx 374 61.47 -----  
 musbjunx 365 61.47 -----  
 humdjunx 41 34.82 t-----g-g--agt  
 humbjunx 182 31.12 a-----gtgg--gc  
 humdjunx 602 31.12 c-----c-ggg-gc  
 humdjunx 602 31.12 c-----c-ggg-gc

PROBE: C:\HITACHI\JUNMIX.PRP  
 HYBRIDIZATION: C:\HITACHI\HUMBJUNX.CDS  
 Length = 377 Hairpin = 2 14  
 Locus Pos Tm  
 humbjunx 377 61.55 -----  
 musbjunx 368 61.55 -----  
 humdjunx 383 28.12 tg-cg-c--g-----  
 musdjunx 383 28.12 tg-ca-c--g-----  
 musdjunx 383 28.12 tg-ca-c--g-----

PROBE: C:\HITACHI\JUNMIX.PRP  
 HYBRIDIZATION: C:\HITACHI\HUMBJUNX.CDS  
 Length = 389 Hairpin = 3 3  
 Locus Pos Tm  
 humbjunx 389 61.7 -----  
 muscjunx 314 56.65 -c-----  
 musbjunx 380 50.85 -----t--g  
 humcjunx 314 49.35 -t-----g-----  
 humdjunx 395 33.85 -----tt-gc--ag  
 musdjunx 395 33.85 -----tt-gc--aa  
 humcjunx 326 32.35 g-ttcgcc-----tg  
 humdjunx 404 32.35 --ttcgcc-----t-  
 muscjunx 326 32.35 gcttcgcc-----tg  
 musdjunx 253 30.85 gacg-gct-ct-----  
 humbjunx 953 30.65 g-----t--c-cagct-  
 musdjunx 83 27.3 cc-gcggg-gt-----g

PROBE: C:\HITACHI\JUNMIX.PRP  
 HYBRIDIZATION: C:\HITACHI\HUMBJUNX.CDS  
 Length = 397 Hairpin = 4 1  
 Locus Pos Tm  
 humbjunx 397 61.55 -----  
 muscjunx 322 53.44 -----g---  
 humcjunx 322 45.33 -----g-----g---  
 musbjunx 388 41.38 -----t--g-----t  
 humdjunx 214 36.83 cccctgc-----  
 humdjunx 99 36.16 cg----gc-c-----  
 musdjunx 261 34.55 -ct-----gatct  
 humdjunx 400 33.27 c---ag-----g---  
 musdjunx 400 33.27 c---ag-----a---  
 humcjunx 334 32.28 -----tgcg--c-  
 humdjunx 412 32.28 -----t-a-g-c-  
 muscjunx 334 32.28 -----tgcg--c-  
 humbjunx 658 30.17 cc-cc-----gt---  
 humdjunx 241 28.95 -c--cacc-c-----  
 humdjunx 342 28.95 c-cca-ca-----ag  
 musbjunx 606 28.95 ---ct-a-ac-----  
 musdjunx 229 28.95 -c-ctgcg-c-----  
 musdjunx 91 26.67 -gt-----gcc-ccg

## FIG. 26A CONT. (2)

PROBE: C:\HITACHI\JUNMIX.PRP  
 HYBRIDIZATION: C:\HITACHI\HUMBJUNX.CDS  
 Length = 417 Hairpin = 2 15  
 Locus Pos Tm  
 humbjunx 417 60.08 -----  
 musbjunx 408 55.52 -----c-----  
 humdjunx 420 37.3 c-----g-----g---t-a-  
 musbjunx 61 29.0 g---gg-----ca-cctgt-  
 muscjunx 672 26.27 gc-gc-----a-g--aga--

PROBE: C:\HITACHI\JUNMIX.PRP  
 HYBRIDIZATION: C:\HITACHI\HUMBJUNX.CDS  
 Length = 461 Hairpin = 4 9  
 Locus Pos Tm  
 humbjunx 461 61.63 -----  
 musbjunx 452 61.63 -----  
 musbjunx 452 61.63 -----

PROBE: C:\HITACHI\JUNMIX.PRP  
 HYBRIDIZATION: C:\HITACHI\HUMBJUNX.CDS  
 Length = 467 Hairpin = 2 13  
 Locus Pos Tm  
 humbjunx 467 61.7 -----  
 musbjunx 458 51.6 -----c-g-  
 humdjunx 32 29.35 tgagcgg-----gcgg-  
 humdjunx 32 29.35 tgagcgg-----gcgg-

PROBE: C:\HITACHI\JUNMIX.PRP  
 HYBRIDIZATION: C:\HITACHI\HUMBJUNX.CDS  
 Length = 477 Hairpin = 2 4  
 Locus Pos Tm  
 humbjunx 477 61.37 -----  
 humdjunx 489 34.93 c-c---cg-----  
 humdjunx 489 34.93 c-c---cg-----

PROBE: C:\HITACHI\JUNMIX.PRP  
 HYBRIDIZATION: C:\HITACHI\HUMBJUNX.CDS  
 Length = 487 Hairpin = 3 3  
 Locus Pos Tm  
 humbjunx 487 61.14 -----  
 musdjunx 74 51.0 ct-----  
 humdjunx 499 45.64 -----t---g  
 humdjunx 527 30.72 cc-c-c-----  
 musdjunx 97 30.72 ttc-c-----g  
 musdjunx 580 30.72 -cc-----t-g  
 musdjunx 637 30.72 cc-cc-----g  
 musdjunx 637 30.72 cc-cc-----g

PROBE: C:\HITACHI\JUNMIX.PRP  
 HYBRIDIZATION: C:\HITACHI\HUMBJUNX.CDS  
 Length = 498 Hairpin = 3 2  
 Locus Pos Tm  
 humbjunx 498 61.26 -----  
 humbjunx 498 61.26 -----

PROBE: C:\HITACHI\JUNMIX.PRP  
 HYBRIDIZATION: C:\HITACHI\HUMBJUNX.CDS  
 Length = 504 Hairpin = 3 2  
 Locus Pos Tm  
 humbjunx 504 61.47 -----  
 musbjunx 495 40.35 c--a-----t-  
 humdjunx 609 35.29 cg-----cgggg-  
 humdjunx 609 35.29 cg-----cgggg-



FIG. 27

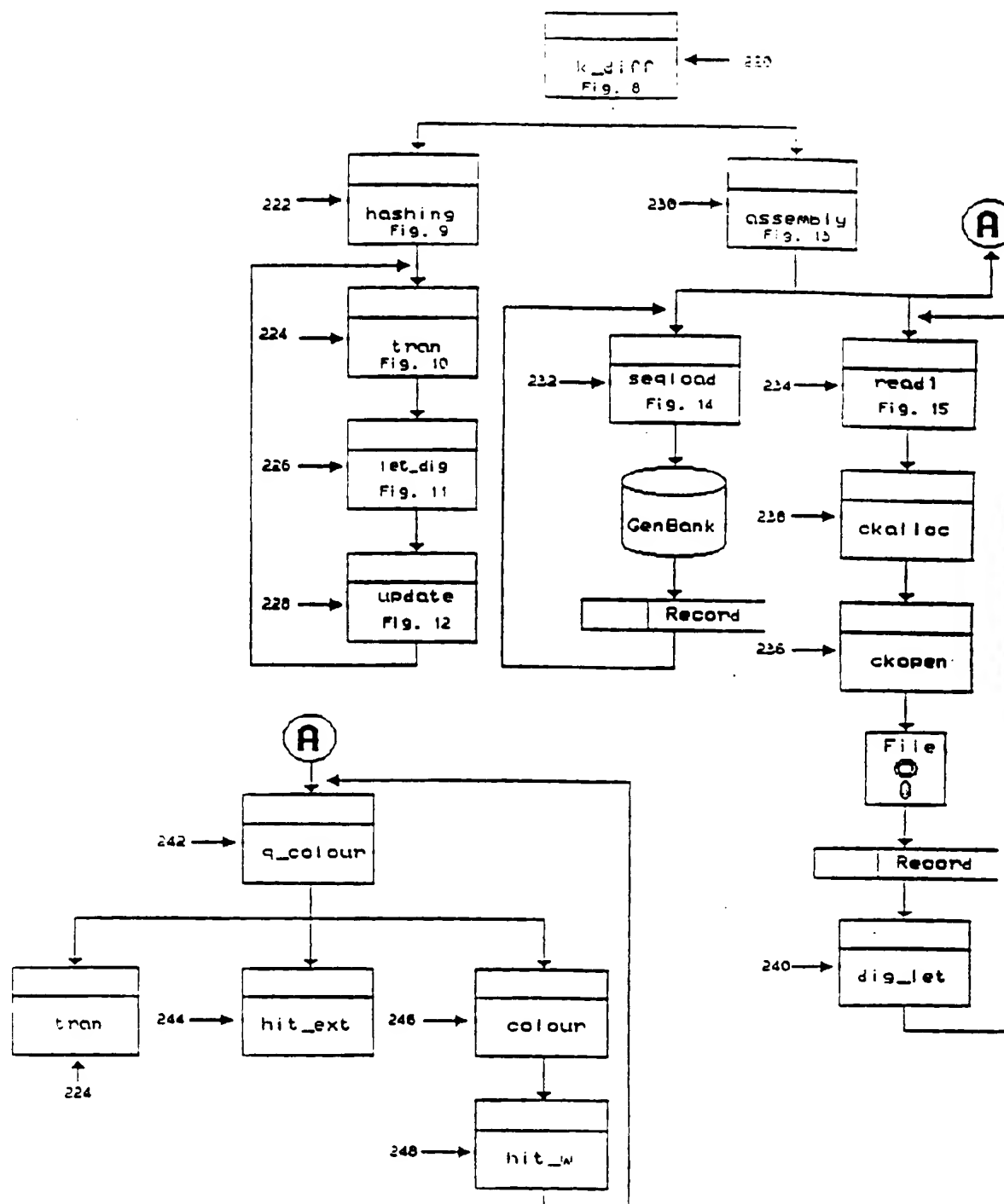


FIG. 28

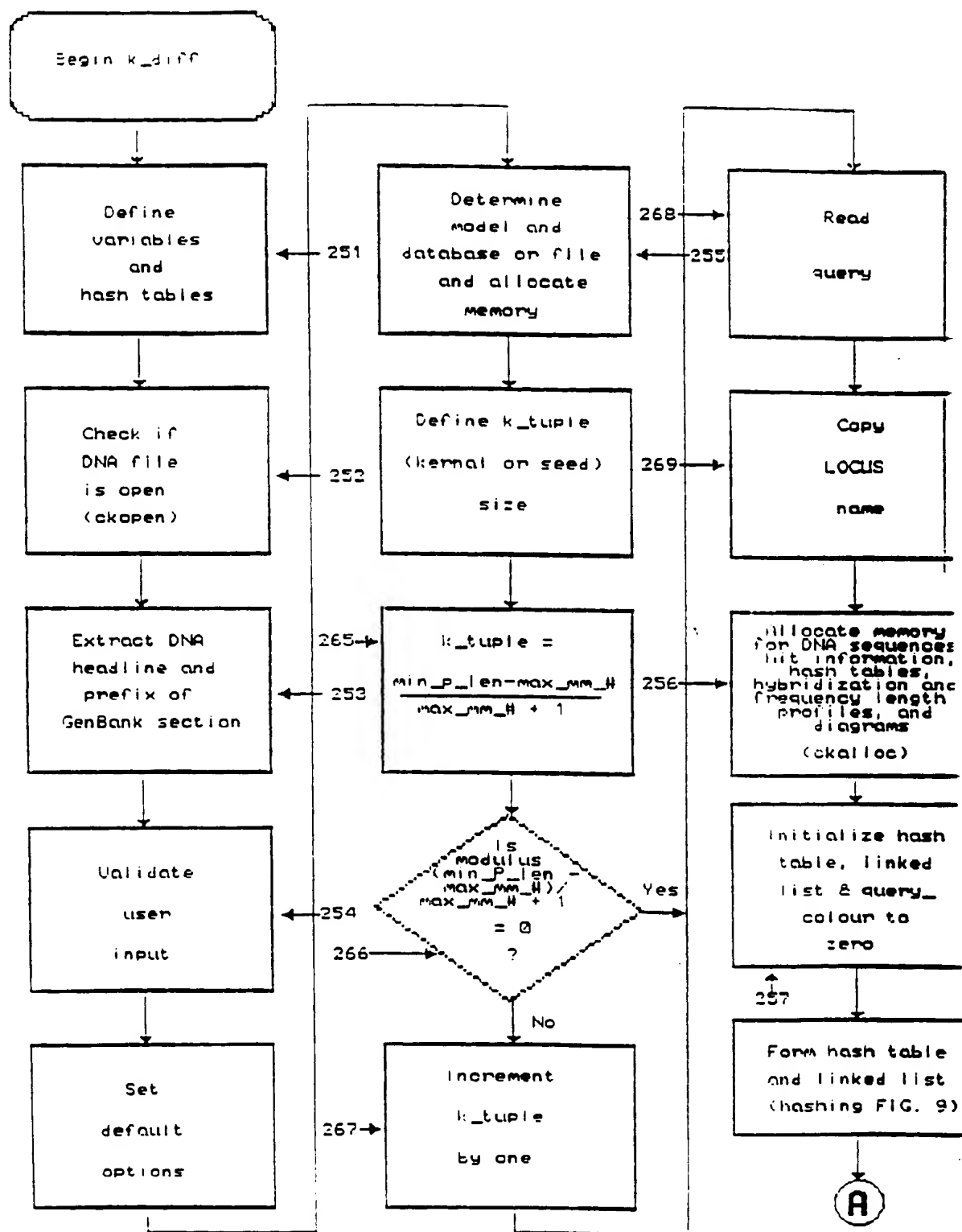
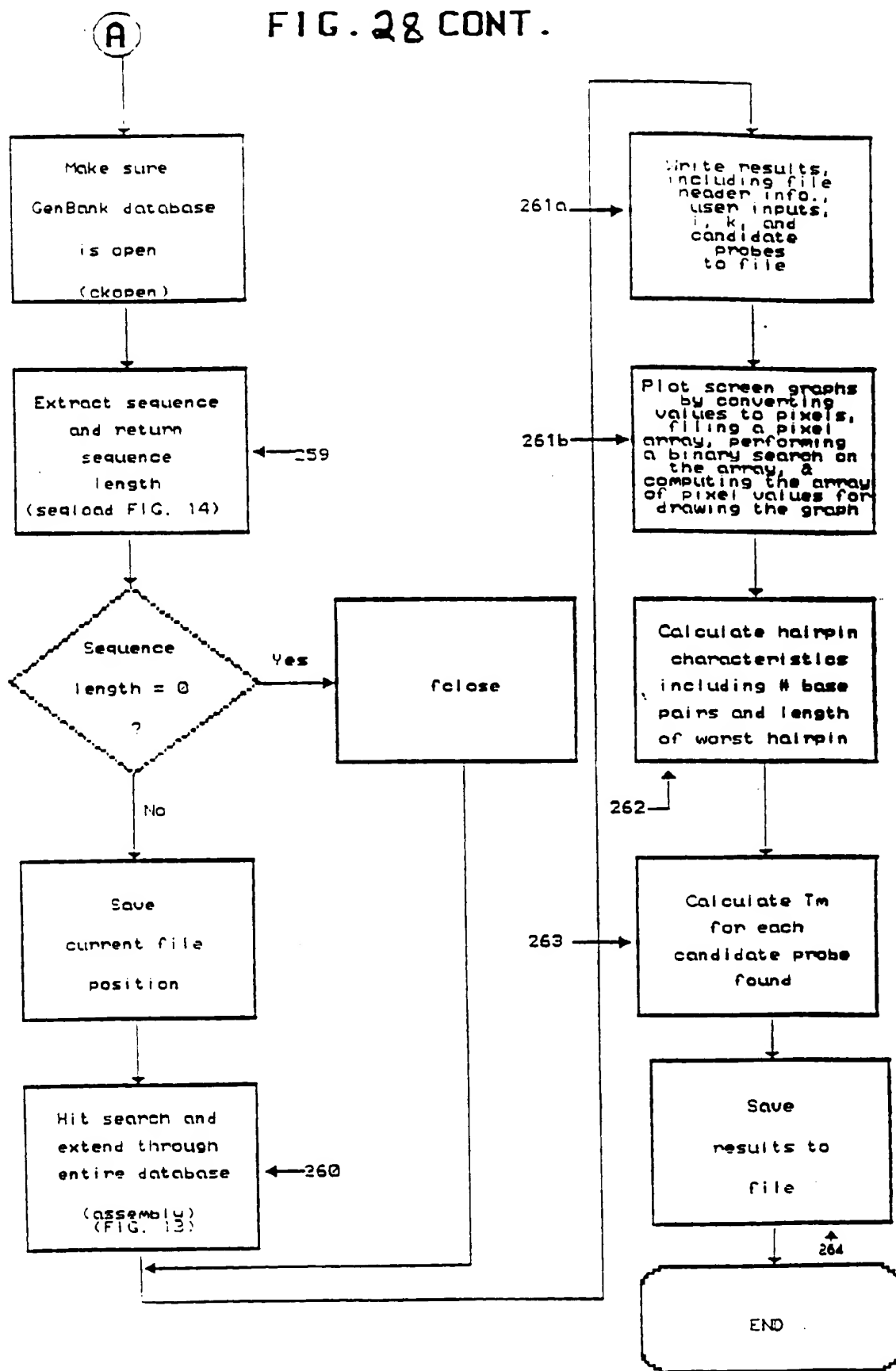


FIG. 28 CONT.



30/59

FIG. 29

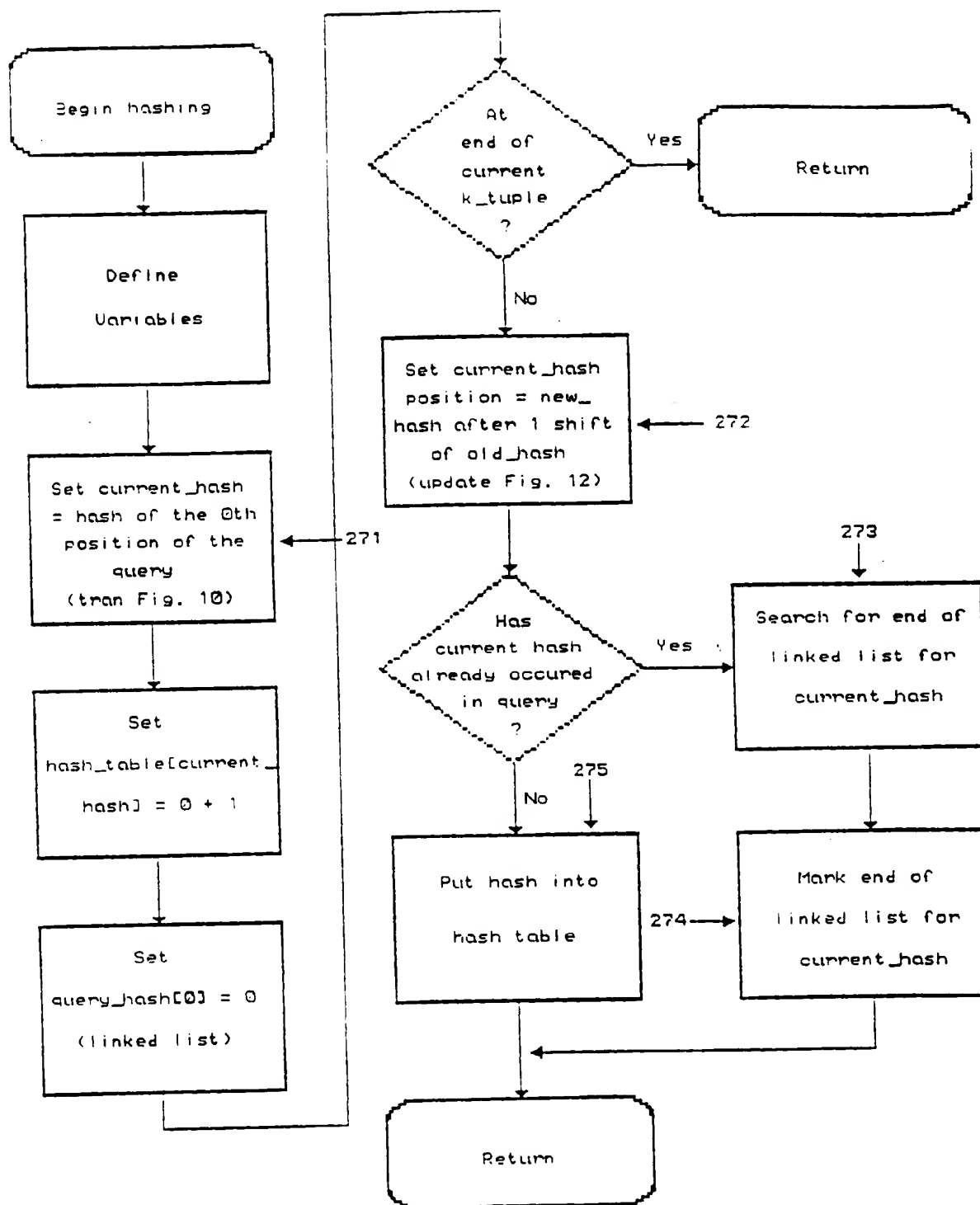
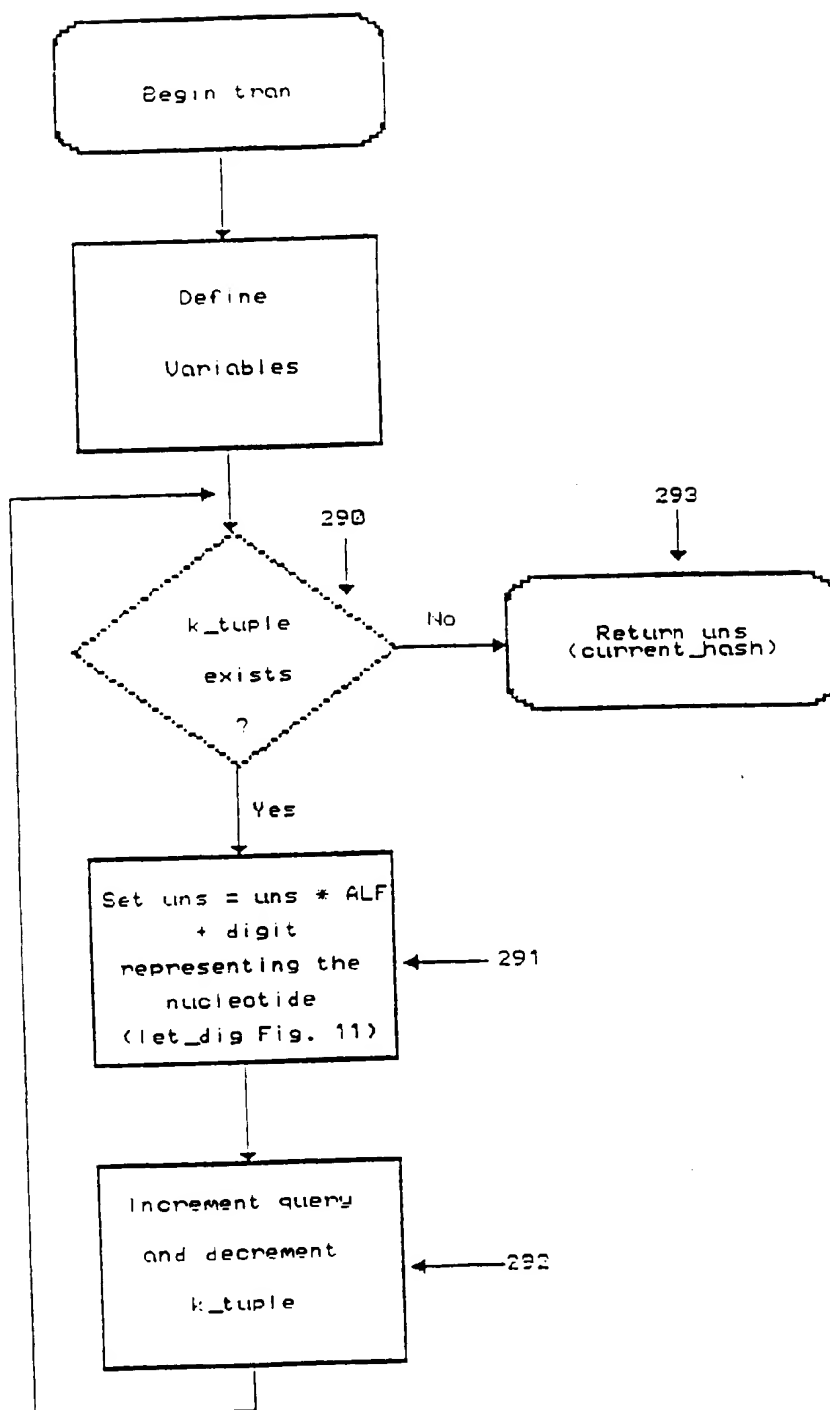
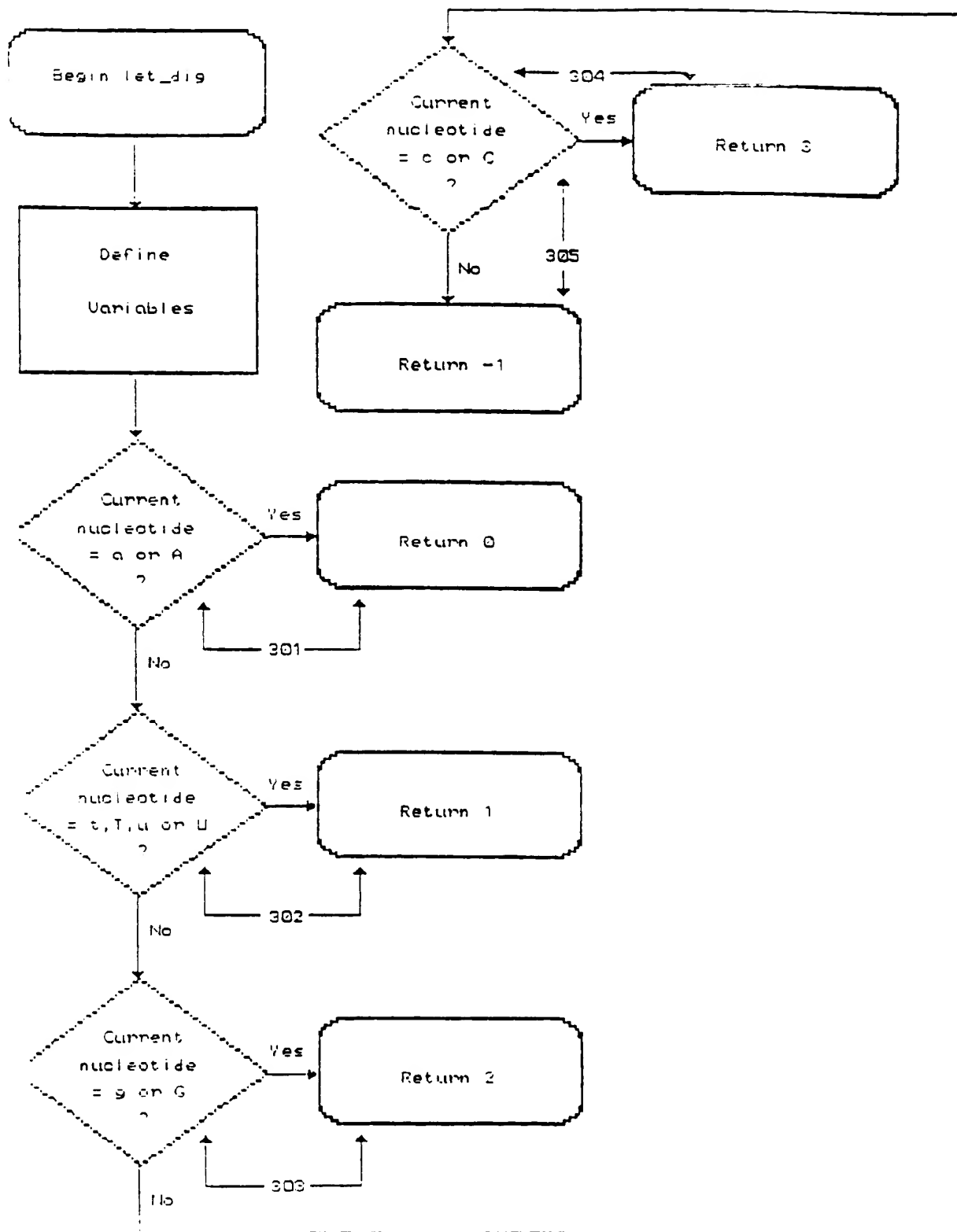


FIG. 30



32/59

FIG. 31



33/59

FIG. 32

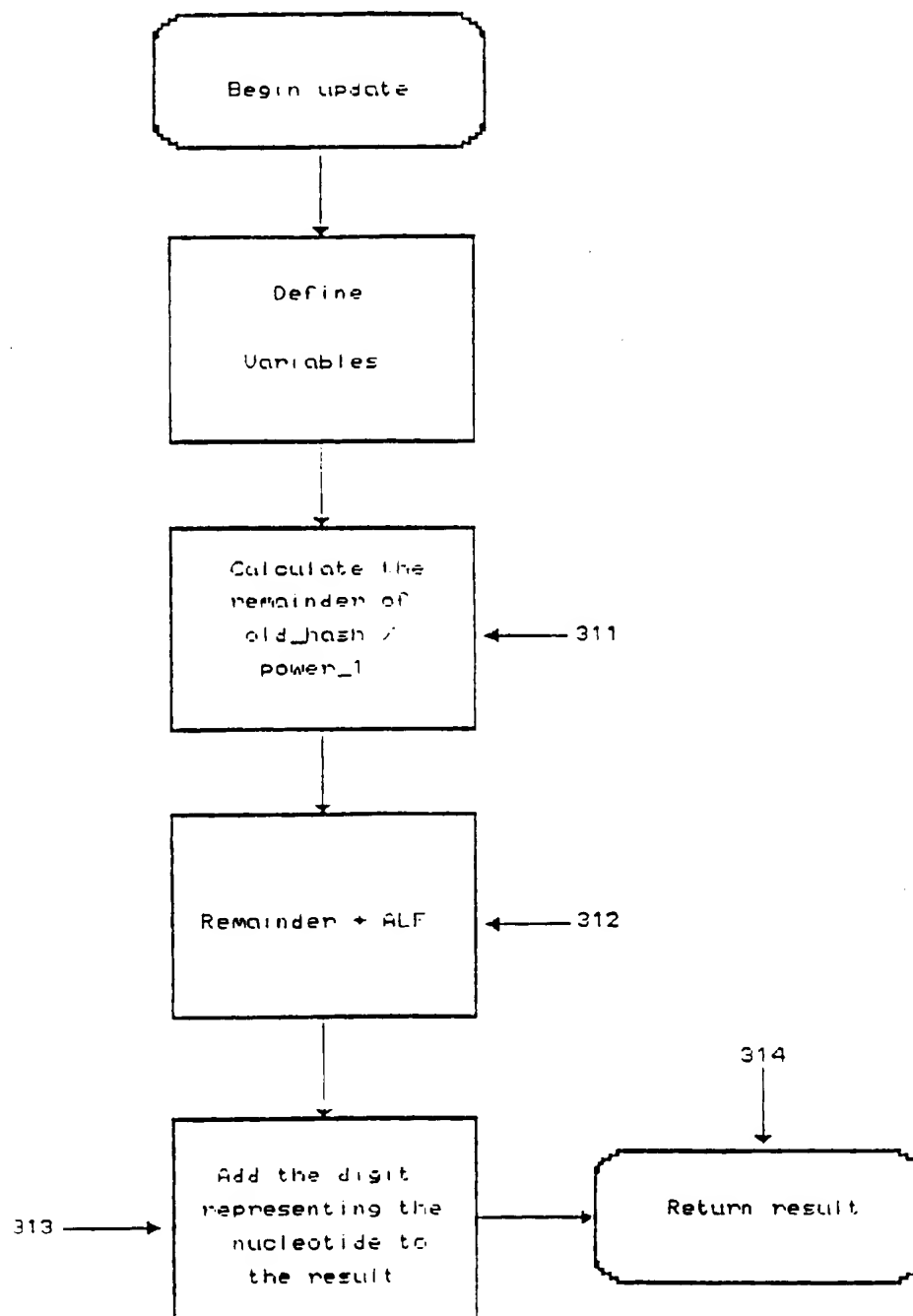


FIG. 33

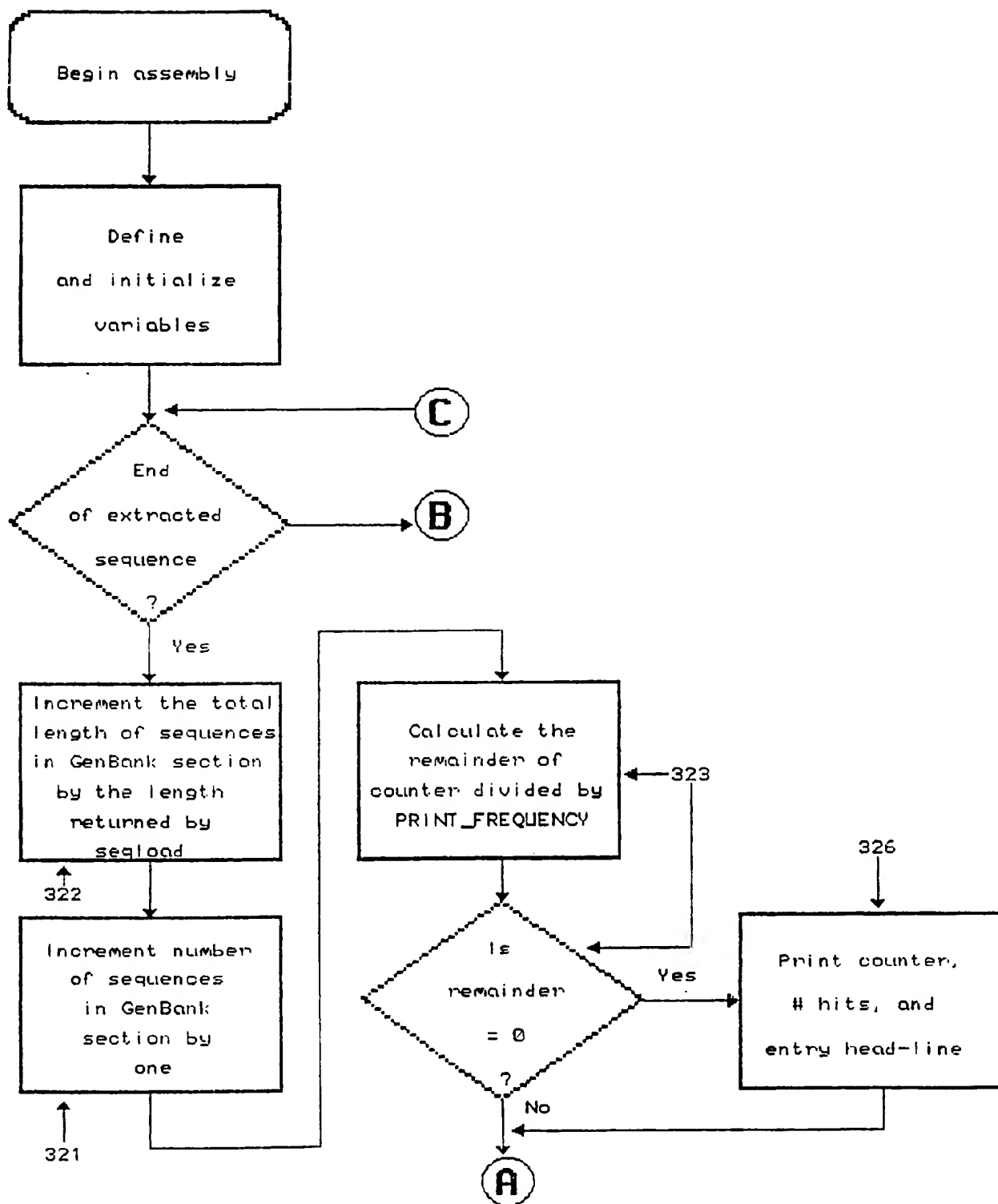
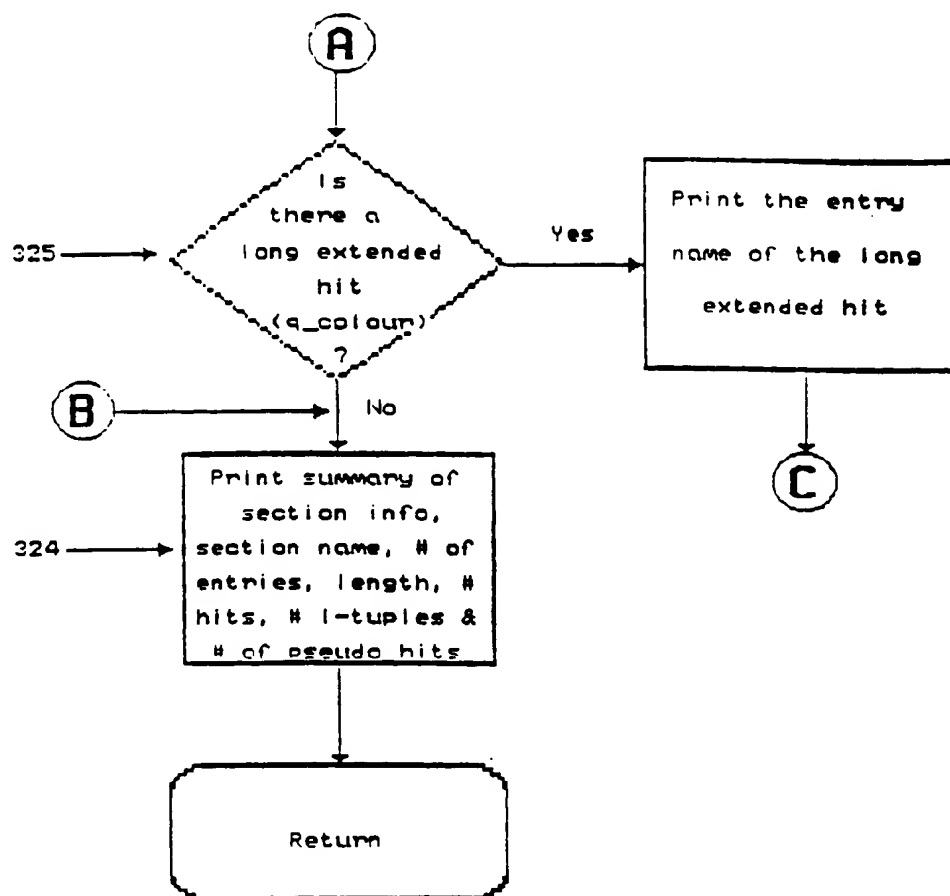




FIG. 33 CONT.



36/59

FIG. 34

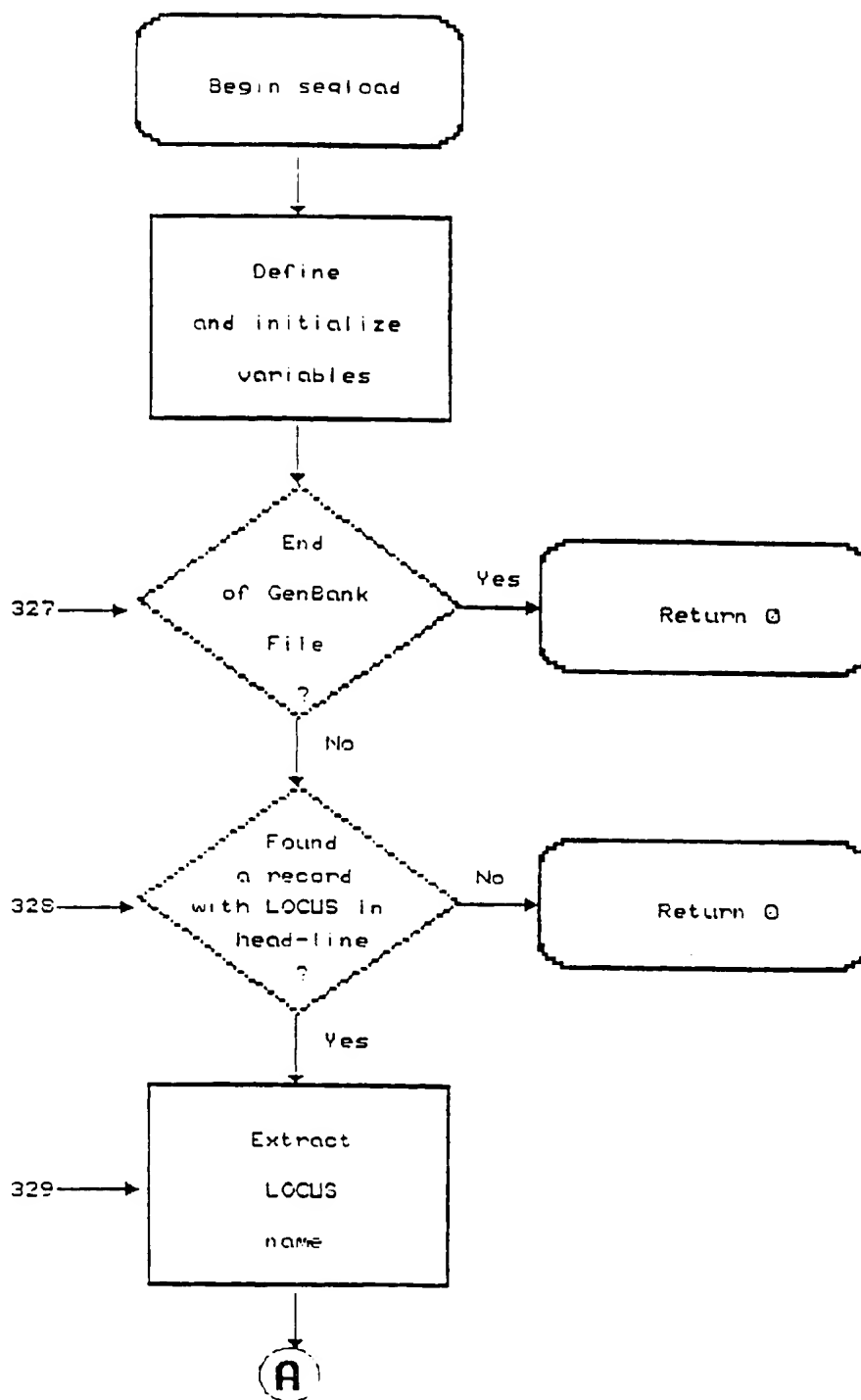


FIG. 34 CONT.

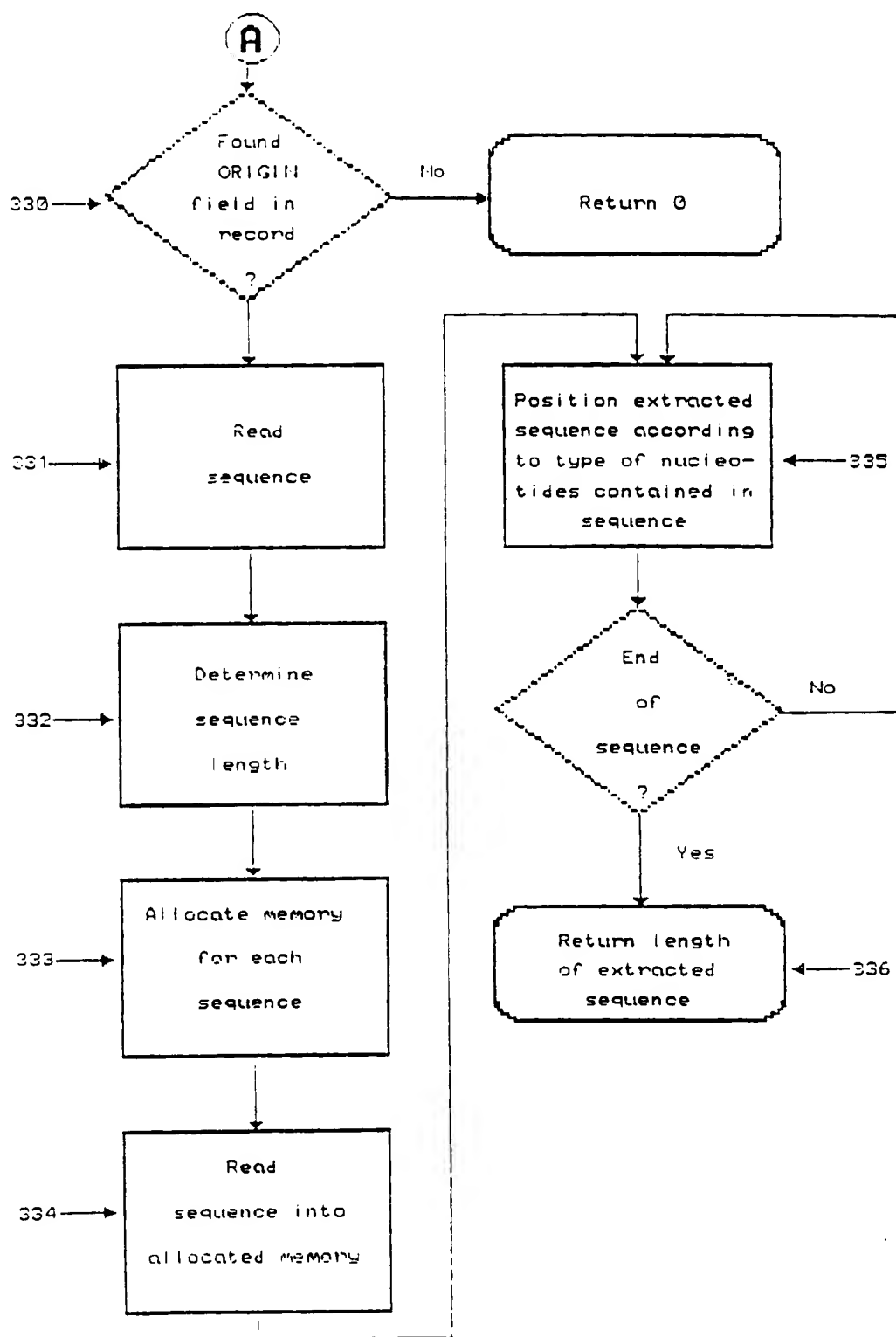


FIG. 35

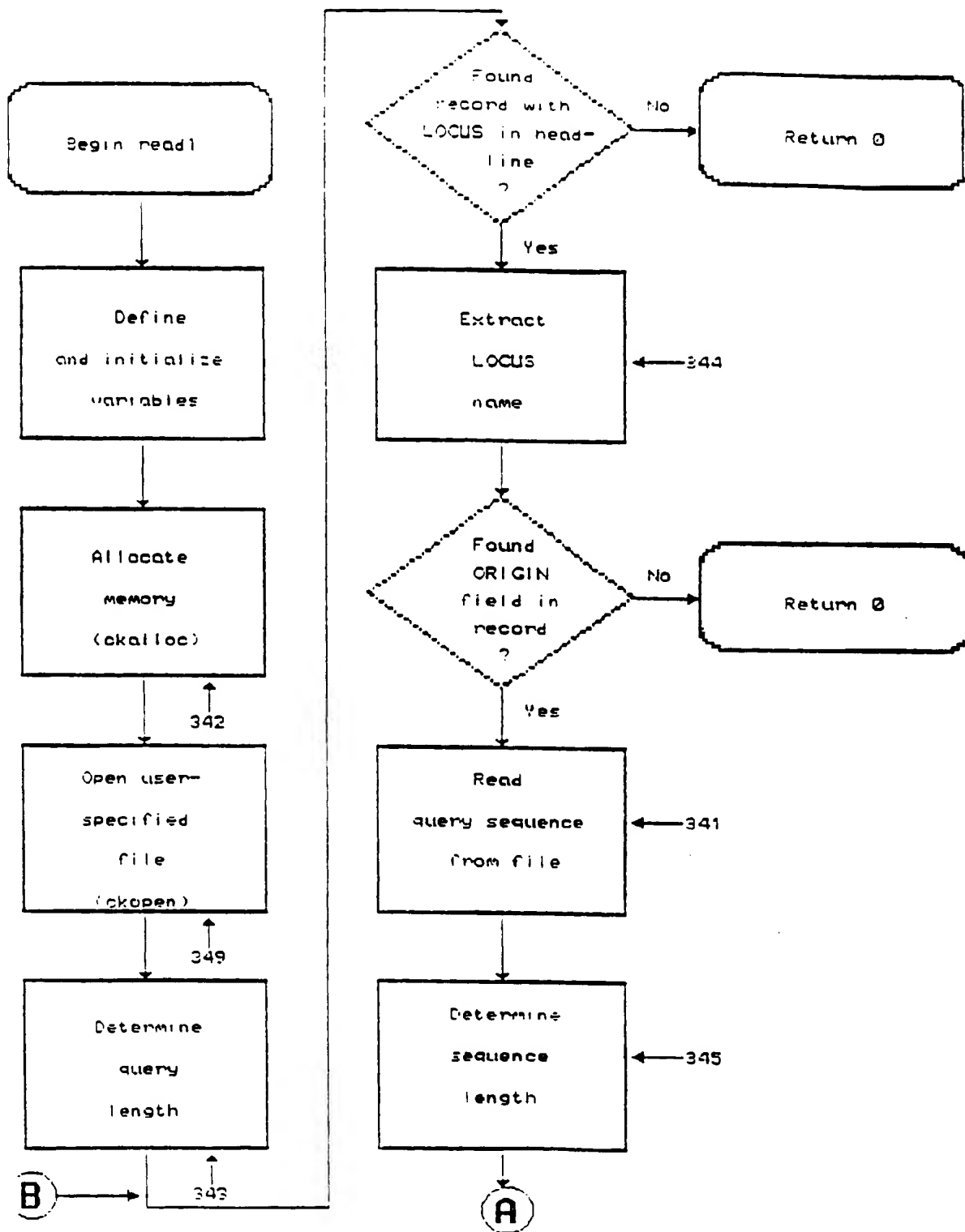


FIG. 35 CONT.

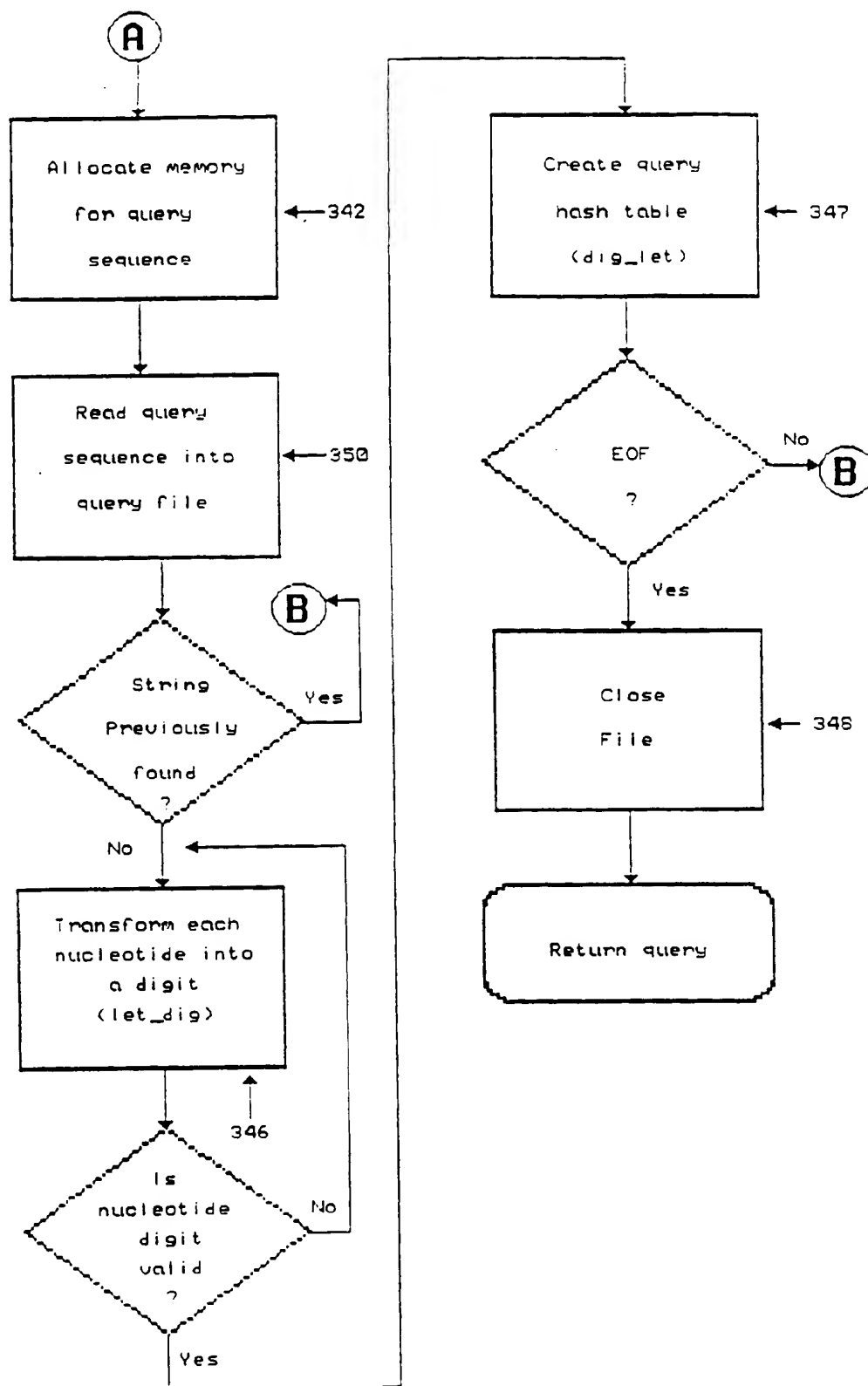


FIG. 36

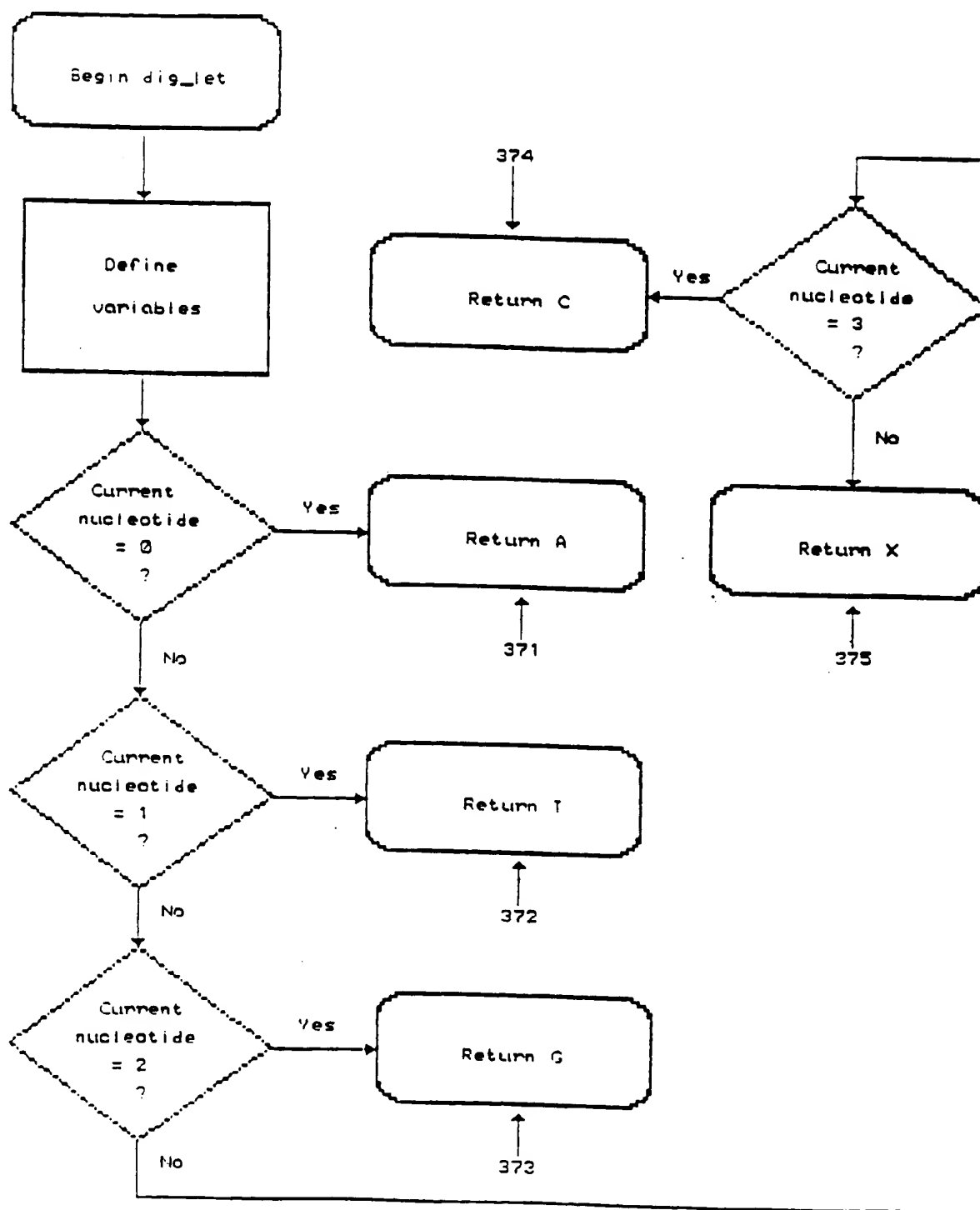
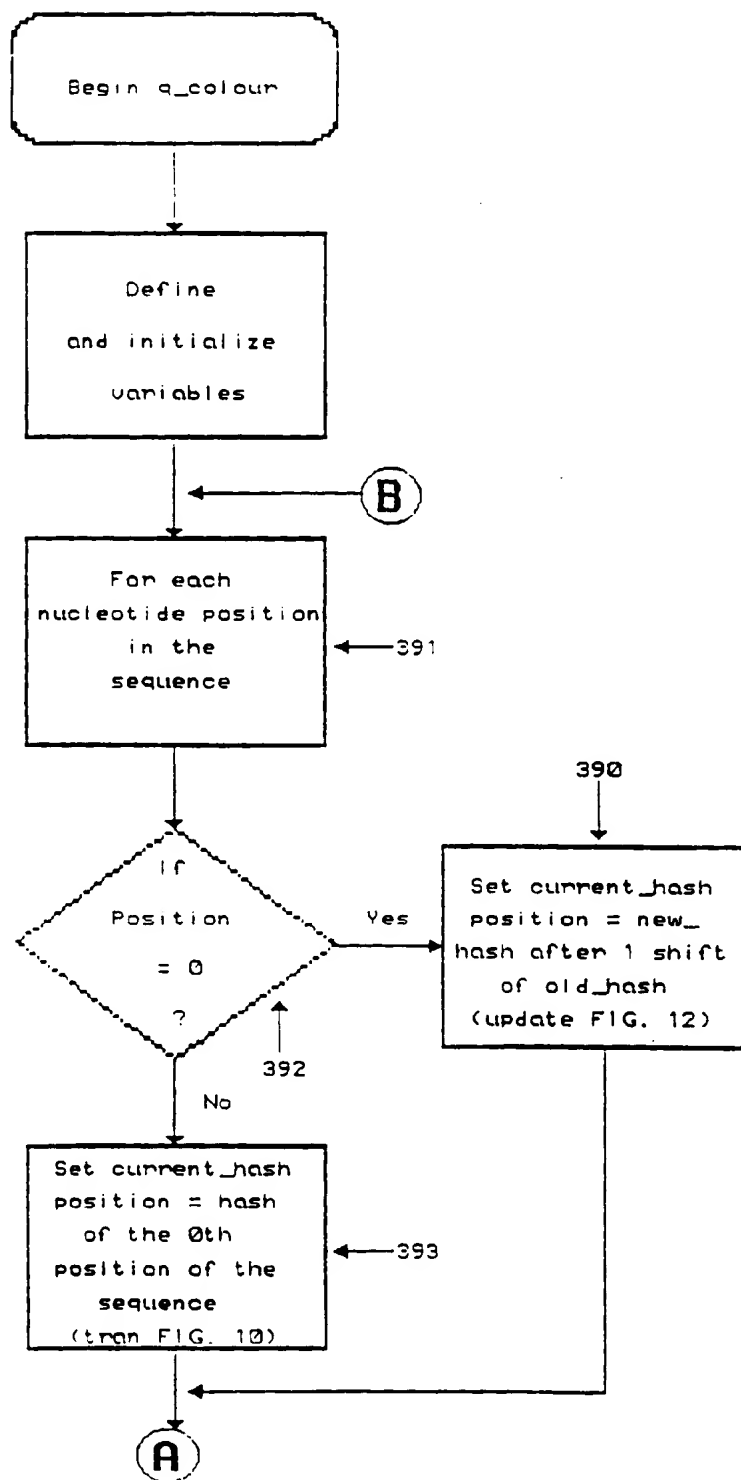
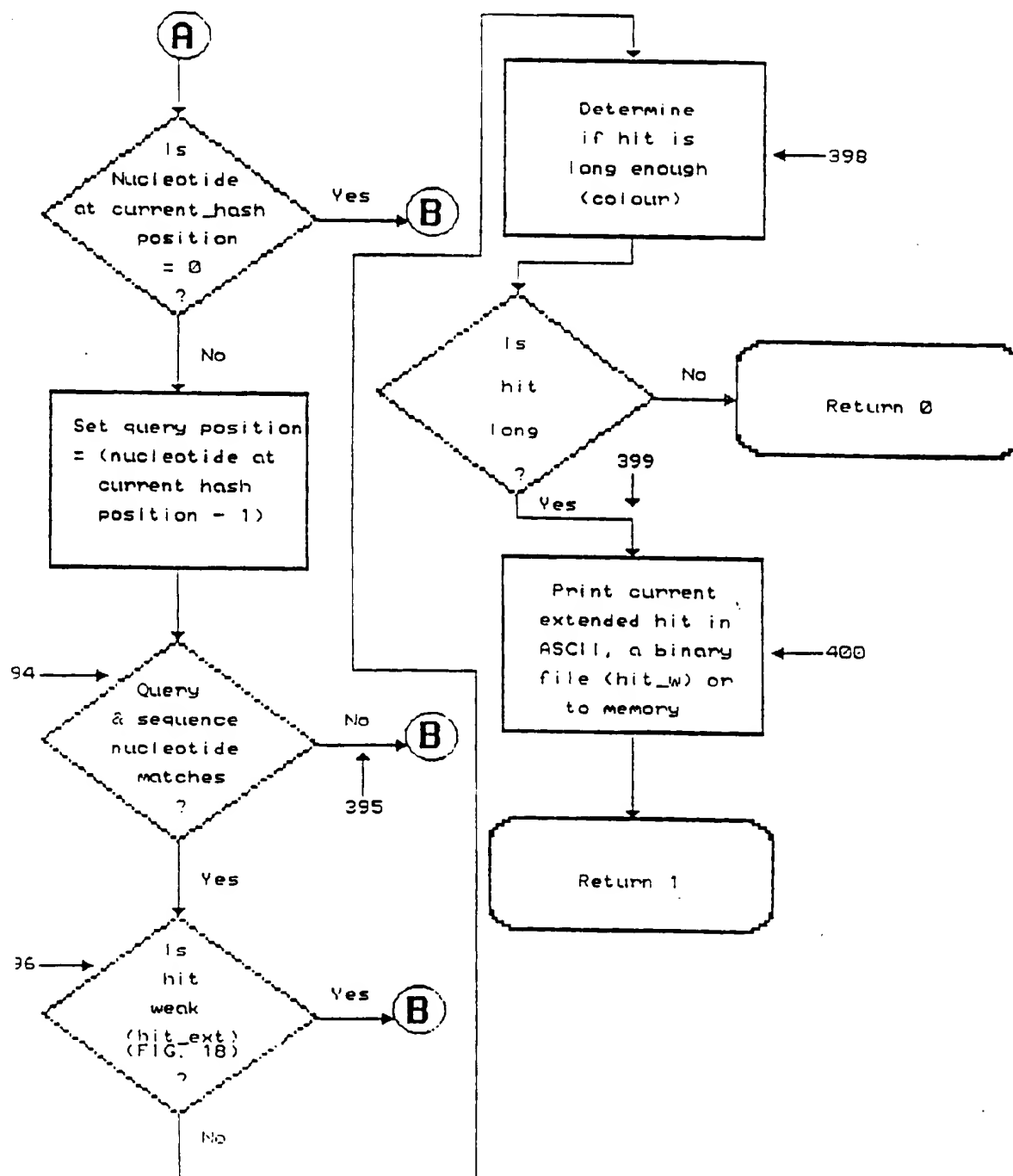


FIG. 37



42/59

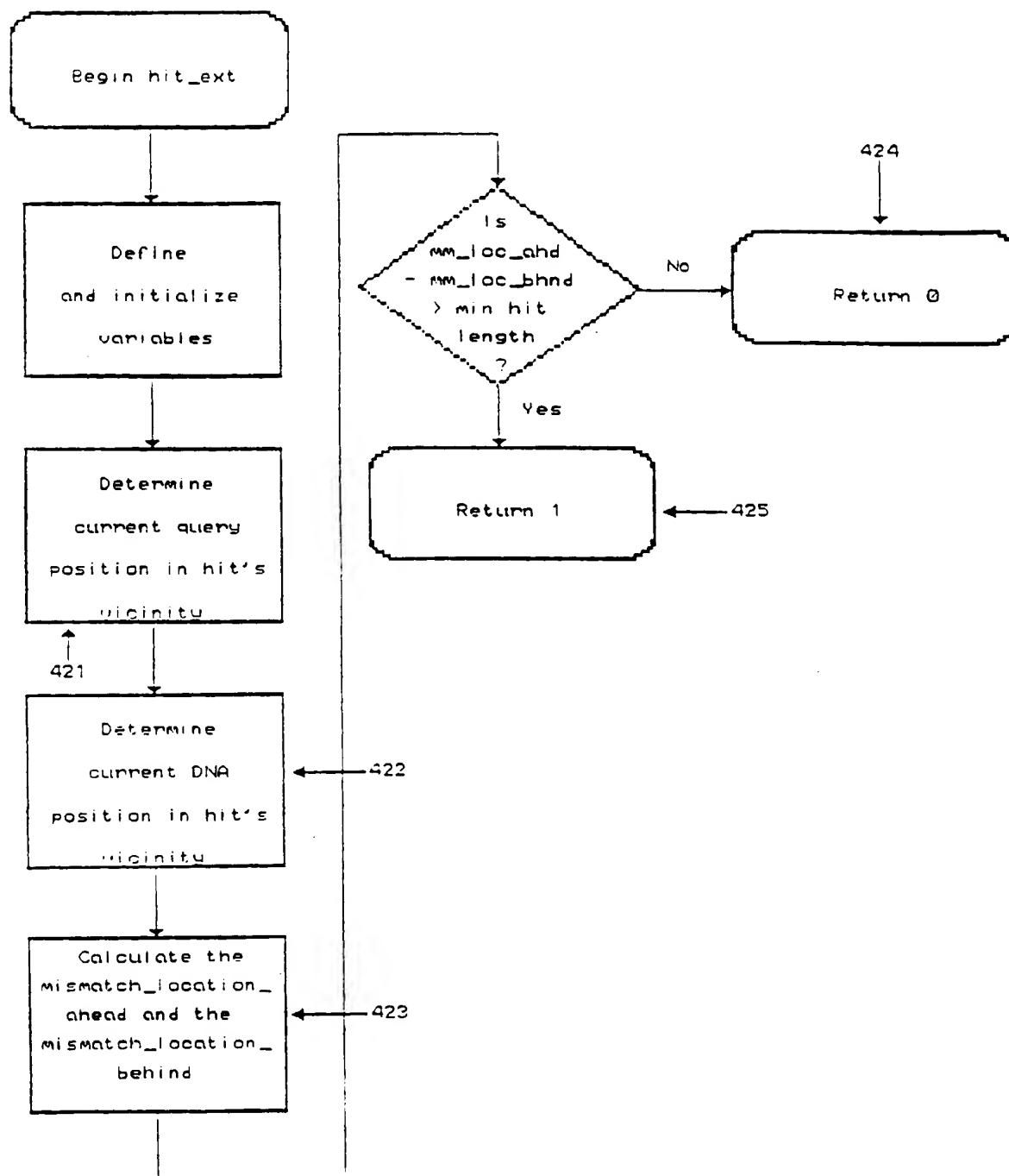
FIG. 37 CONT.





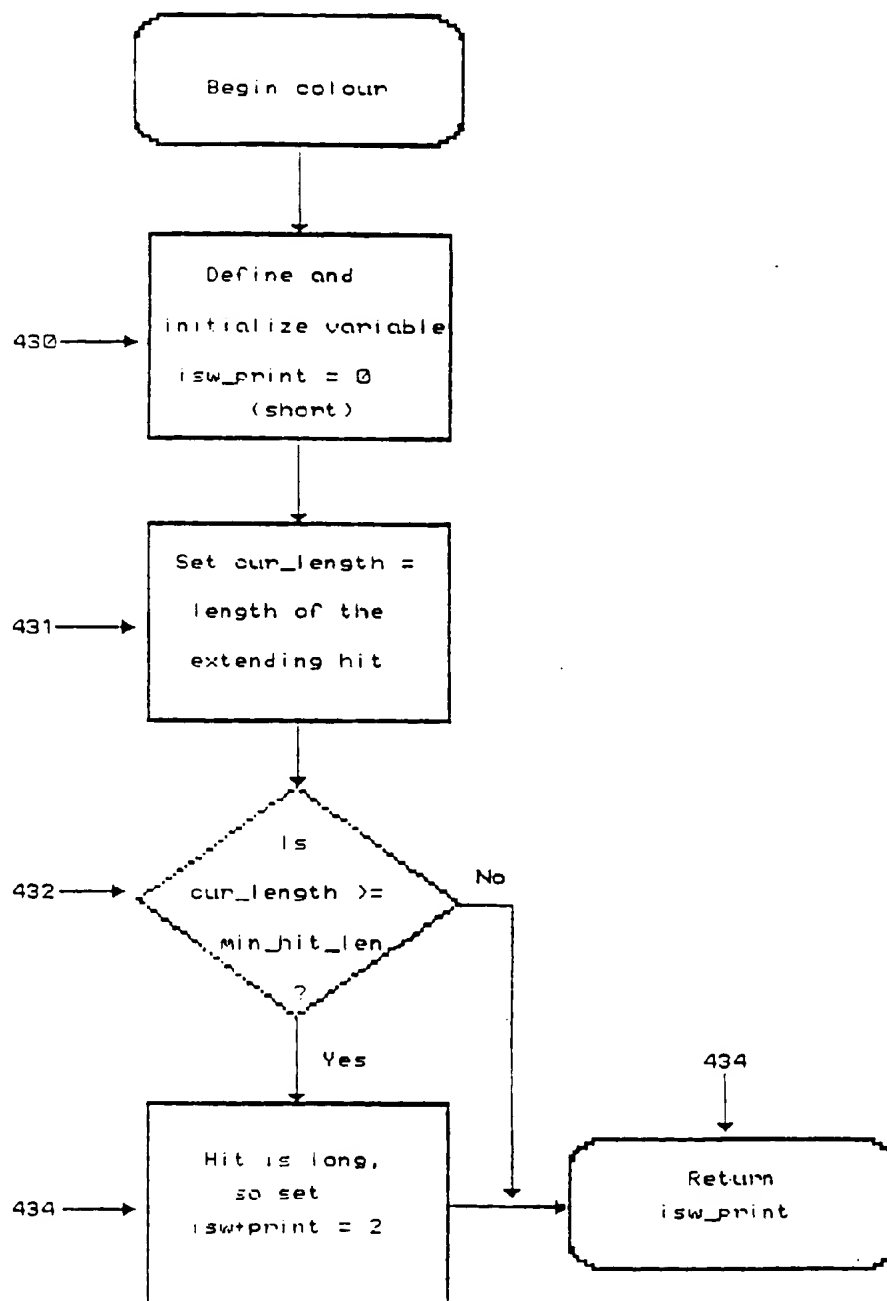
43/59

FIG. 38



44/59

FIG. 39



45/59

FIG. 40

## OligoProbe DesignStation

Probes: C:\HITACHI\HUMBJUNX.CDS  
 Datatbase: C:\HITACHI\JUNMIX.SEO

Mismatch Model, l = 21, k = 4

Position length	Mismatches					screensN					Probe
	0	1	2	3	4	5	6	7	8		
1 21	1	1	1	1	1	1	1	1	1	1	ATGTGCACTAAAATGGAACAG
2 21	1	1	1	1	1	1	1	1	1	1	TGTGCACTAAAATGGAACAGC
3 21	1	1	1	1	1	1	1	1	1	1	GTGCACTAAAATGGAACAGCC
4 21	1	1	1	1	1	1	1	1	1	1	TGCACTAAAATGGAACAGCCC
5 21	1	1	1	1	1	1	1	1	1	1	GCACTAAAATGGAACAGCCCT
6 21	1	1	1	1	1	1	1	1	1	1	CACTAAAATGGAACAGCCCTT
7 21	1	1	1	1	1	1	1	1	1	1	ACTAAAATGGAACAGCCCTTC
8 21	1	1	1	1	1	1	1	1	1	1	CTAAAATGGAACAGCCCTTCT
9 21	1	1	1	1	1	1	1	1	1	1	TAAAATGGAACAGCCCTTCTA
10 21	1	1	1	1	1	1	1	1	1	1	AAAATGGAACAGCCCTTCTAC
11 21	1	1	1	1	1	1	1	1	1	1	AAATGGAACAGCCCTTCTACC
12 21	1	1	1	1	1	1	1	1	1	1	AATGGAACAGCCCTTCTACCA
13 21	1	1	1	1	1	1	1	1	1	1	ATGGAACAGCCCTTCTACCAC
14 21	1	1	1	1	1	1	1	1	1	1	TGGAACAGCCCTTCTACCACG
15 21	1	1	1	1	1	1	1	1	1	1	GGAACAGCCCTTCTACCACGA
16 21	1	1	1	1	1	1	1	1	1	1	GAACAGCCCTTCTACCACGAC
17 21	1	1	1	1	1	1	1	1	1	1	AACAGCCCTTCTACCACGACG
18 21	1	1	1	1	1	1	1	1	1	1	ACAGCCCTTCTACCACGACGA
19 21	1	1	1	1	1	1	1	1	1	1	CAGCCCTTCTACCACGACGAC
20 21	1	1	1	1	1	1	1	1	1	1	AGCCCTTCTACCACGACGACT
21 21	1	1	1	1	1	1	1	1	1	1	GCCCTTCTACCACGACGACTC
22 21	1	1	1	1	1	1	1	1	1	1	CCCTTCTACCACGACGACTCA
23 21	1	1	1	1	1	1	1	1	1	1	CCTTCTACCACGACGACTCAT
24 21	1	1	1	1	1	1	1	1	1	1	CTTCTACCACGACGACTCATA
25 21	1	1	1	1	1	1	1	1	1	1	TTCTACCACGACGACTCATAAC
26 21	1	1	1	1	1	1	1	1	1	1	TCTACCACGACGACTCATAACA
27 21	1	1	1	1	1	1	1	1	1	1	CTACCACGACGACTCATAACAC
28 21	1	1	1	1	1	1	1	1	1	1	TACCACGACGACTCATAACACA
29 21	1	1	1	1	1	1	1	1	1	1	ACCACGACGACTCATAACACAG
30 21	1	1	1	1	1	1	1	1	1	1	CCACGACGACTCATAACACAGC
31 21	1	1	1	1	1	1	1	1	1	1	CACGACGACTCATAACACAGCT
32 21	1	1	1	1	1	1	1	1	1	1	ACGACGACTCATAACACAGCTA
33 21	1	1	1	1	1	1	1	1	1	1	CGACGACTCATAACACAGCTAC
34 21	1	1	1	1	1	1	1	1	1	1	GACGACTCATAACACAGCTACG
35 21	1	1	1	1	1	1	1	1	1	1	ACGACTCATAACACAGCTACGG
36 21	1	1	1	1	1	1	1	1	1	1	CGACTCATAACACAGCTACGGG
37 21	1	1	1	1	1	1	1	1	1	1	GACTCATAACACAGCTACGGGA
38 21	1	1	1	1	1	1	1	1	1	1	ACTCATAACACAGCTACGGGAT
39 21	1	1	1	1	1	1	1	1	1	1	CTCATAACACAGCTACGGGATA
40 21	1	1	1	1	1	1	1	1	1	1	TCATAACACAGCTACGGGATAC
41 21	1	1	1	1	1	1	1	1	1	1	CATAACACAGCTACGGGATACG
42 21	1	1	1	1	1	1	1	1	1	1	ATACACAGCTACGGGATACGG
43 21	1	1	1	1	1	1	1	1	1	1	TACACAGCTACGGGATACGGC
44 21	1	1	1	1	1	1	1	1	1	1	ACACAGCTACGGGATACGGCCC
45 21	1	1	1	1	1	1	1	1	1	1	CACAGCTACGGGATACGGCCCG
46 21	1	1	1	1	1	1	1	1	1	1	ACAGCTACGGGATACGGCCCGG
47 21	1	1	1	1	1	1	1	1	1	1	CAGCTACGGGATACGGCCCGGG
48 21	1	1	1	1	1	1	1	1	1	1	AGCTACGGGATACGGCCCGGGC
49 21	1	1	1	1	1	1	1	1	1	1	GCTACGGGATACGGCCCGGGCC
50 21	1	1	1	1	1	1	1	1	1	1	CTACGGGATACGGCCCGGGCCCC
51 21	1	1	1	1	1	1	1	1	1	1	TACGGGATACGGCCCGGGCCCCC
52 21	1	1	1	1	1	1	1	1	1	1	ACGGGATACGGCCCGGGCCCCCT
53 21	1	1	1	1	1	1	1	1	1	1	CGGGATACGGCCCGGGCCCCCTG

FIG. 41

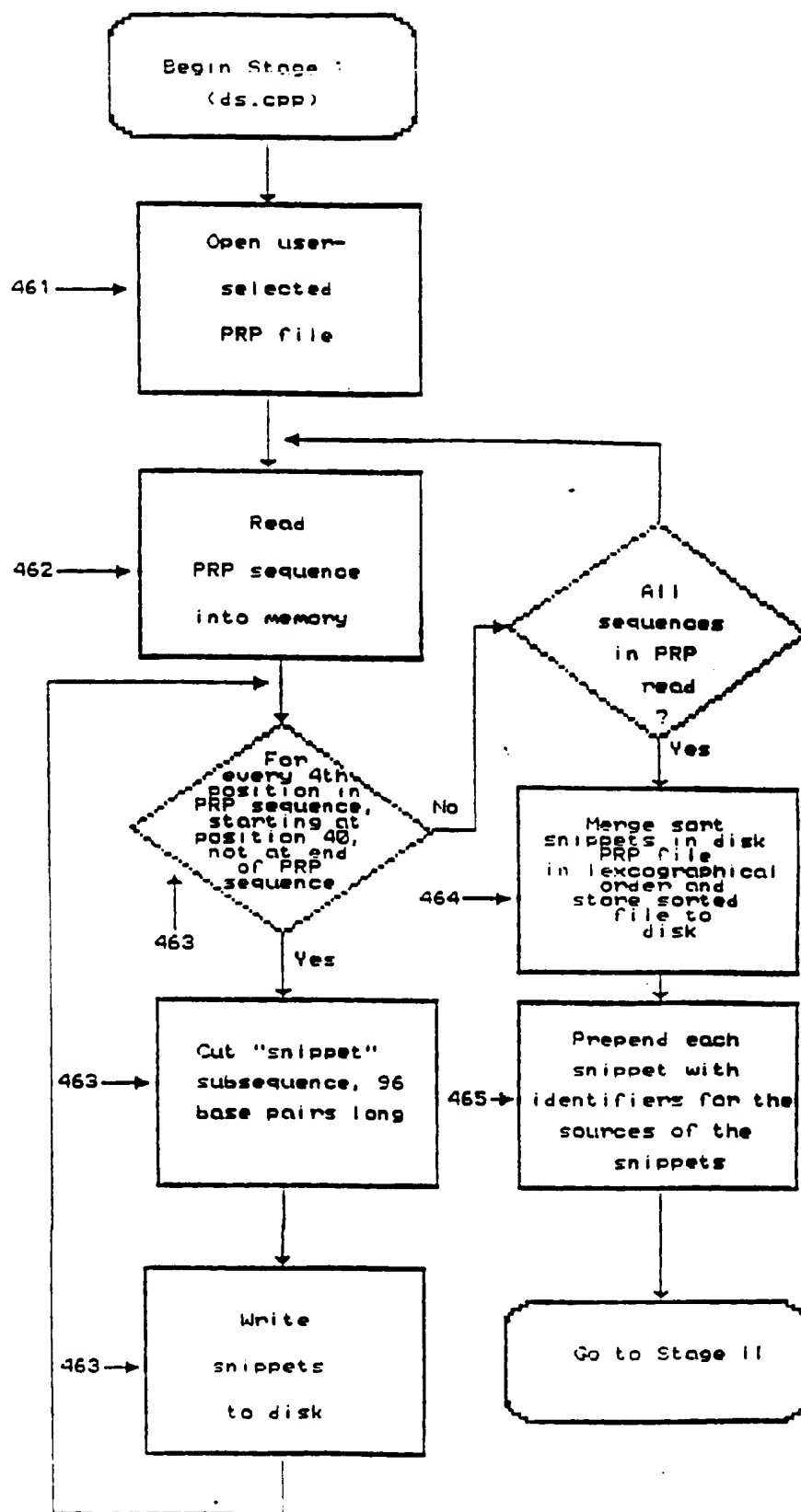


FIG. 42

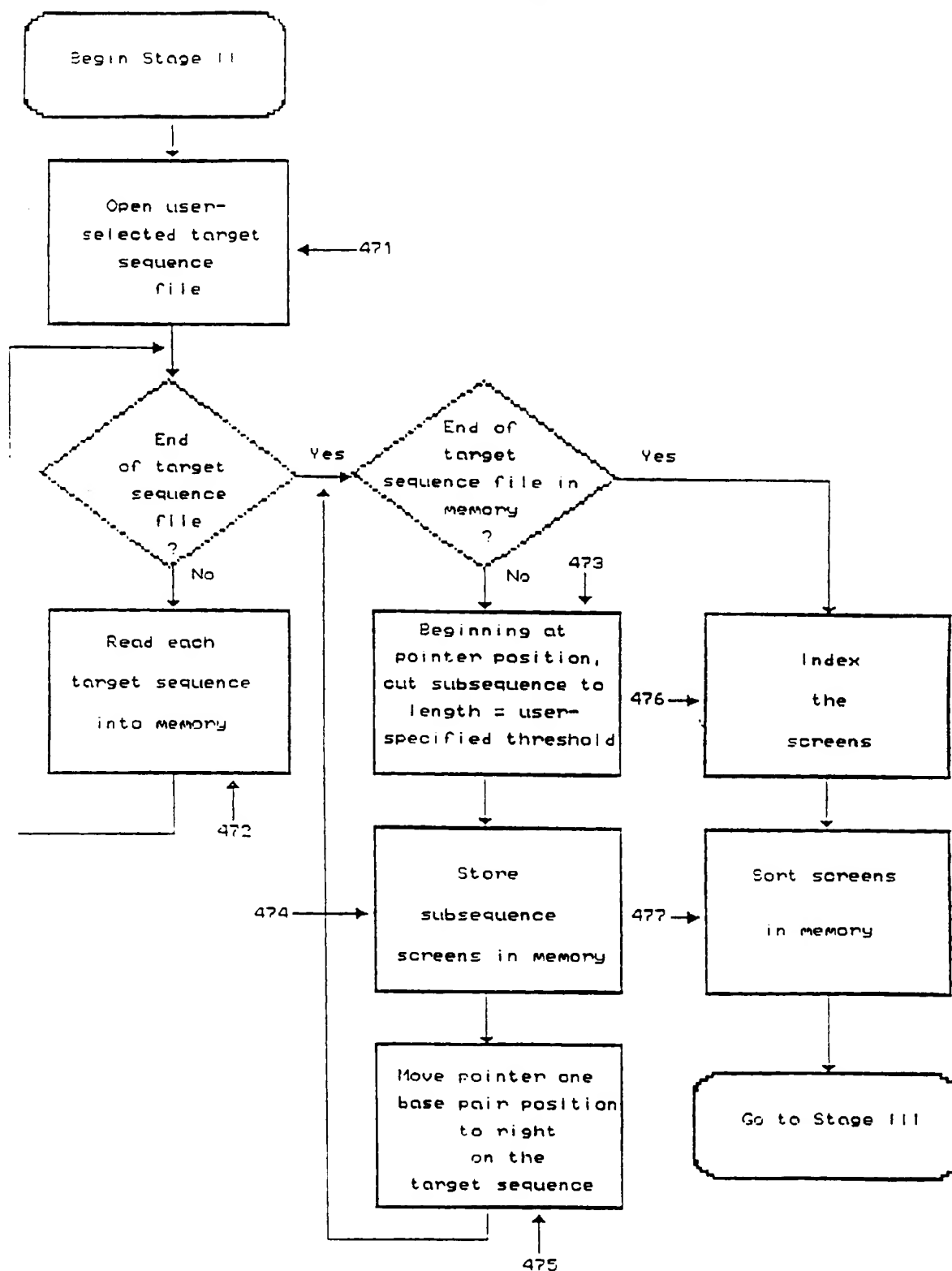


FIG. 43

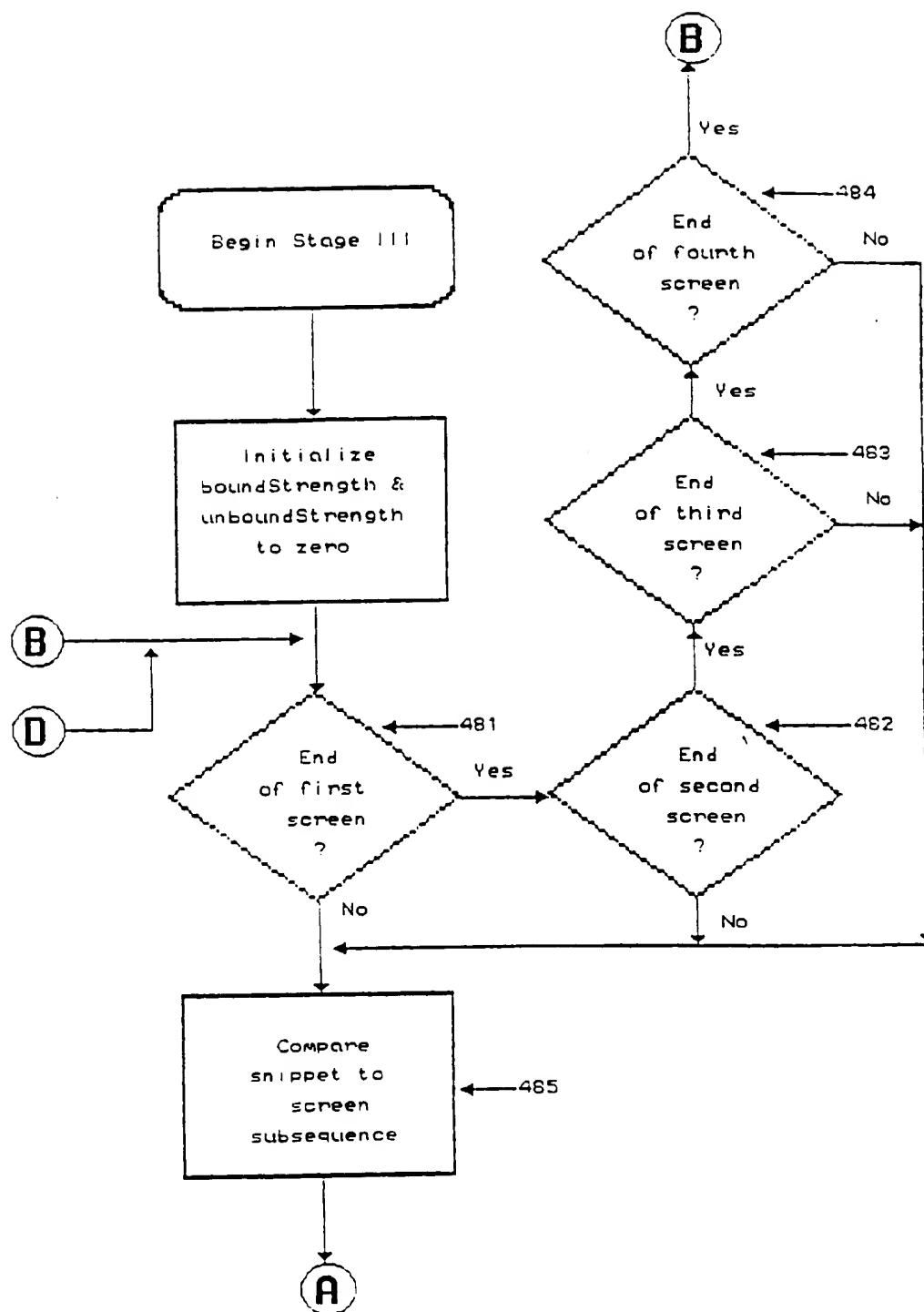


FIG. 43 CONT. 2

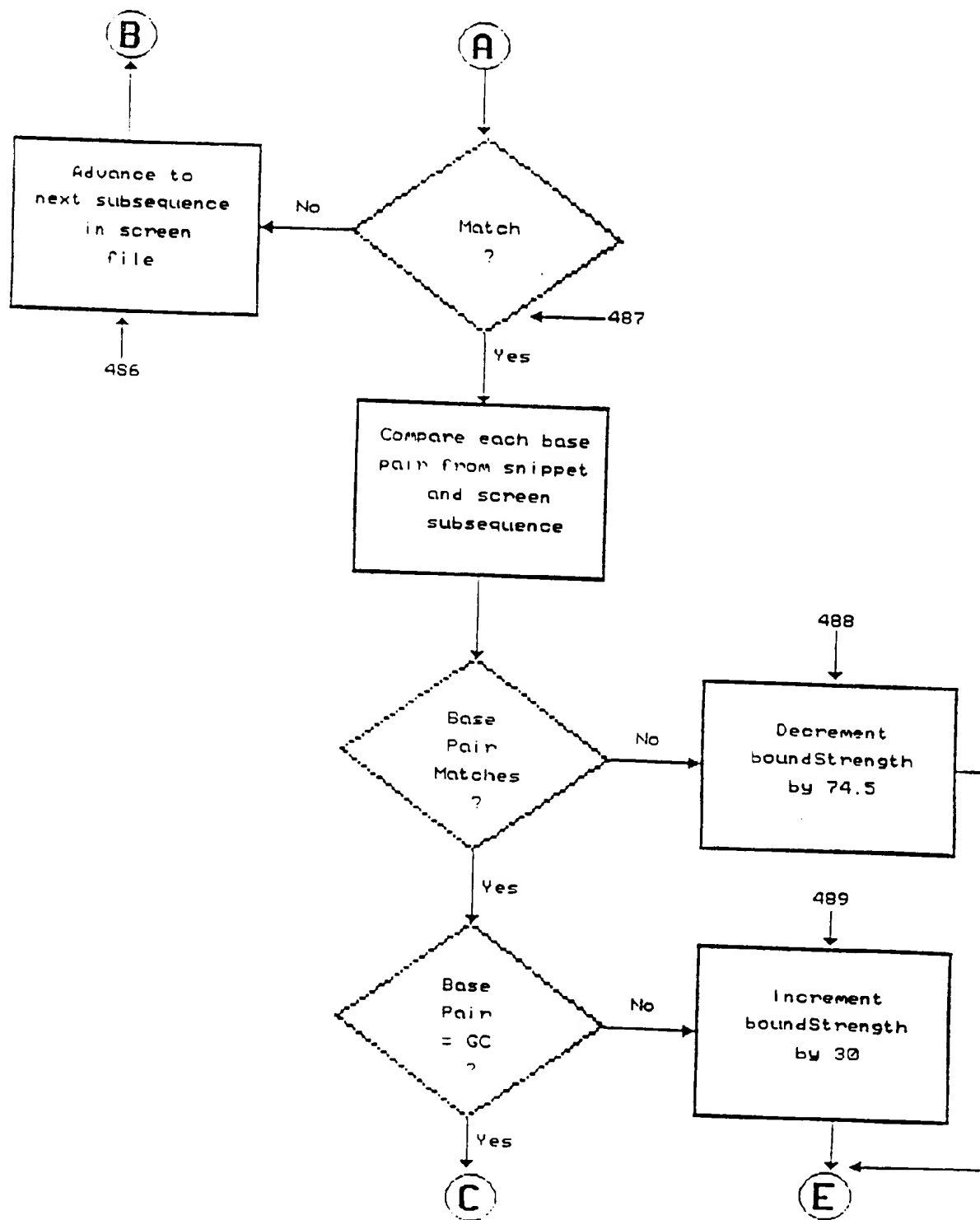


FIG. 43 CONT. 3

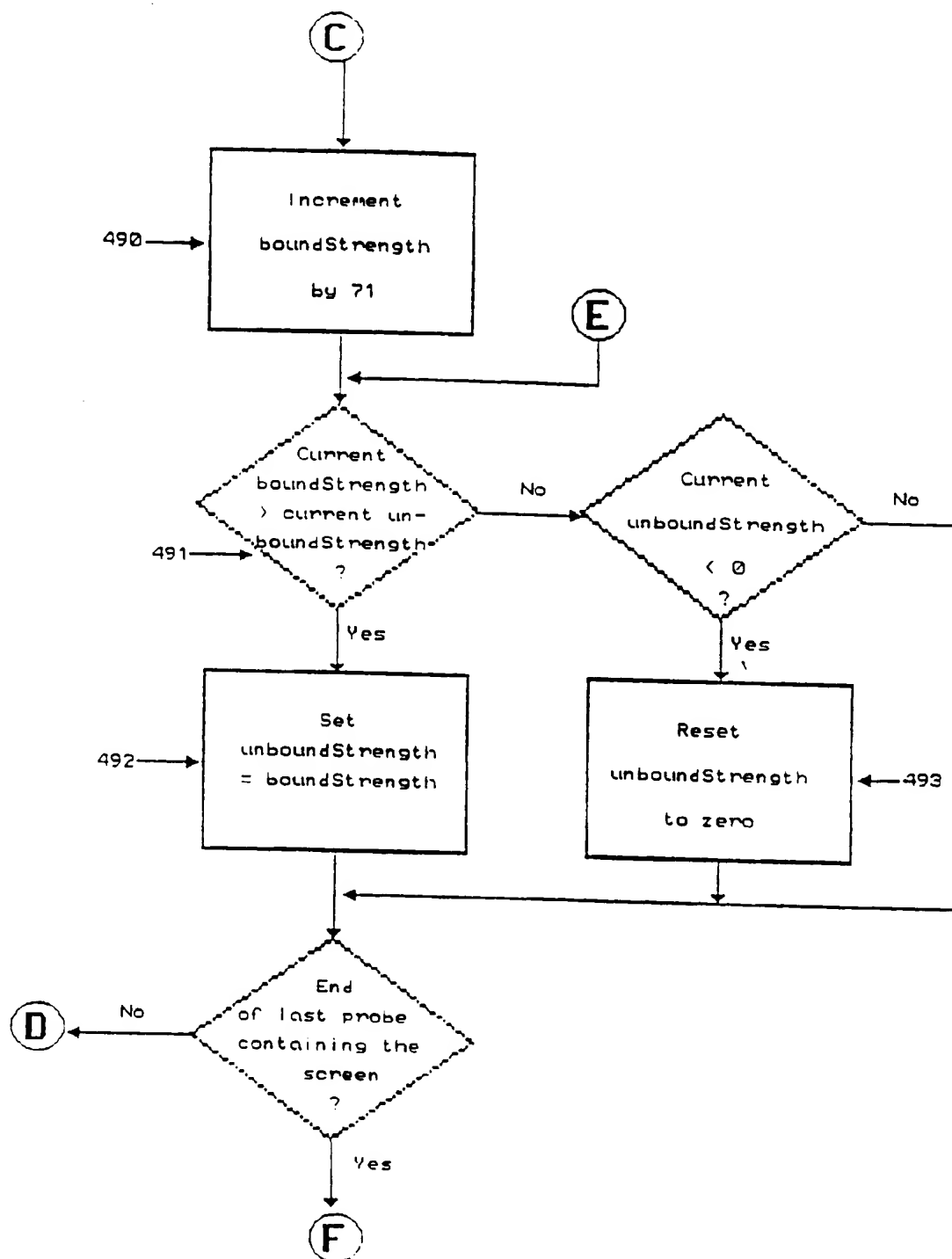
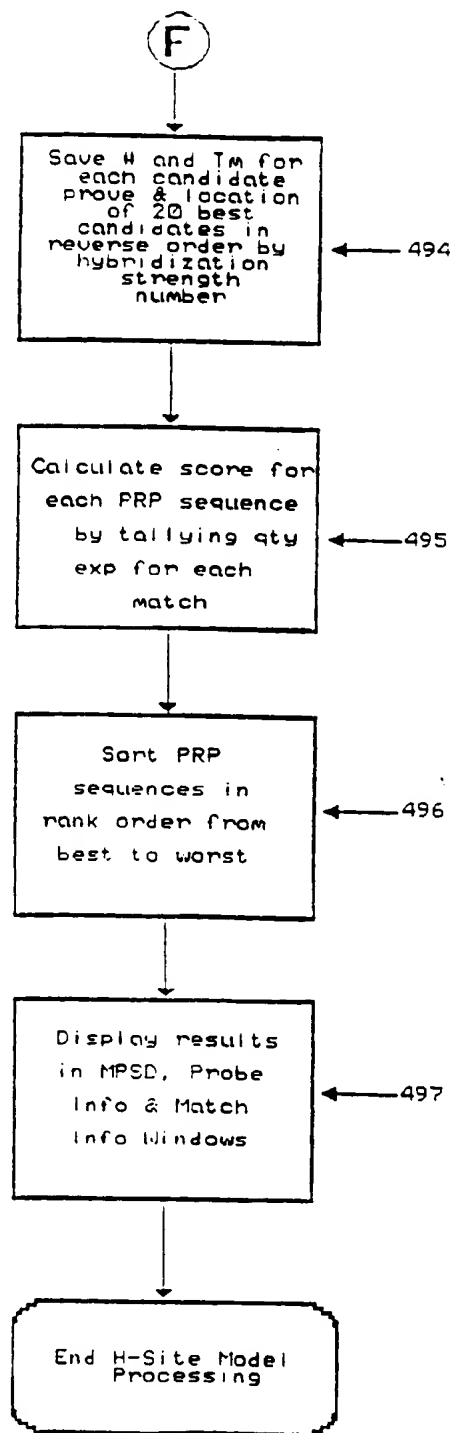




FIG. 43 CONT. 4





5.3/59

FIG. 44B

(Partial File -- 10 pages of 190 pages)

## OligoProbe DesignStation

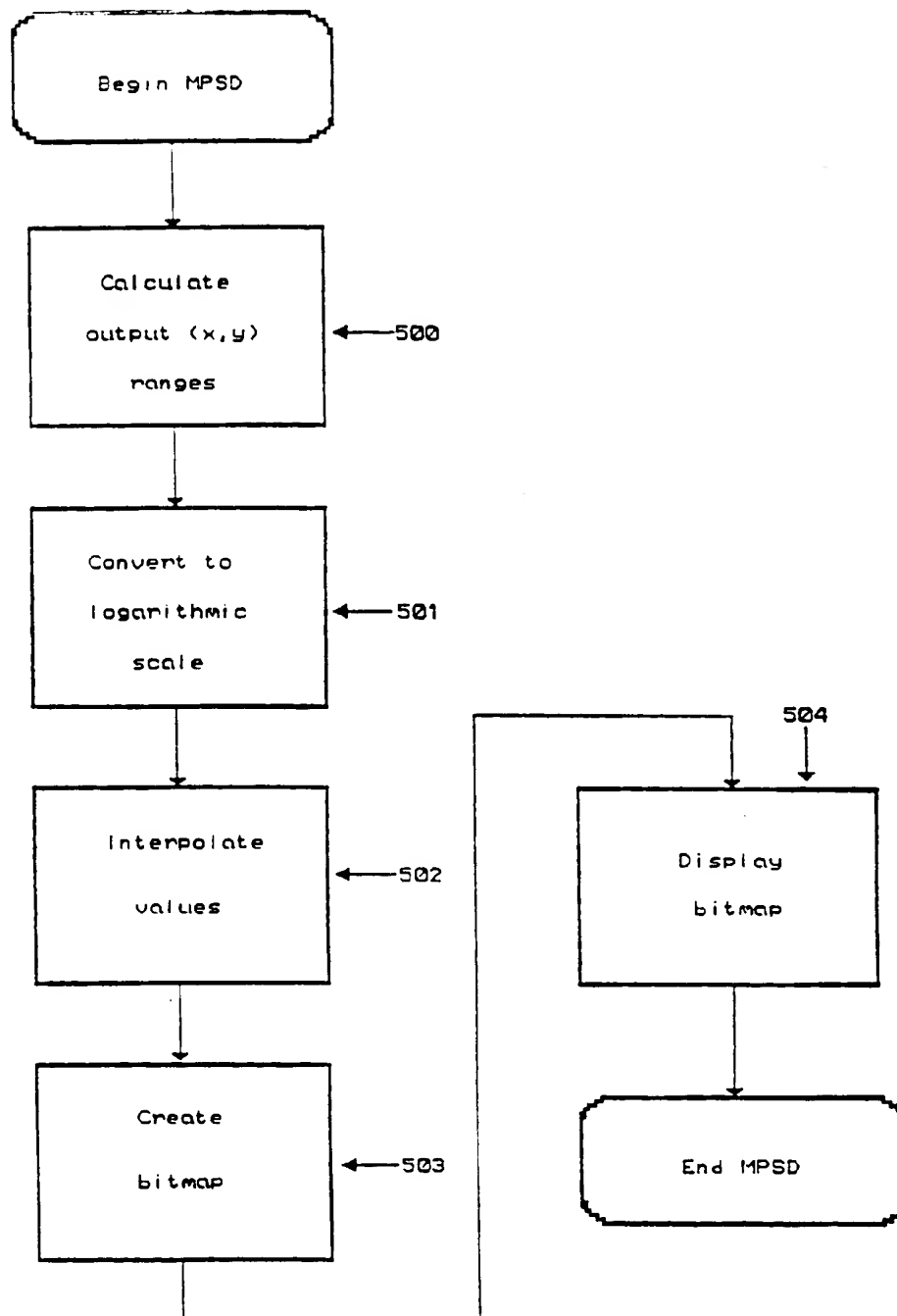
Probes: C:\HITACHI\HUMBJUNX.CDS

Preparation: C:\HITACHI\JUNMIX.PRP

Locus pos	Tm	Locus pos	Tm	Locus pos	Tm
atgtgcactaaaatggaacagcccttctac					
1 30	1 1 2	2 2 2 2 3 4			
humbjunx 1	60.76				
musbjunx 1	50.03				
muscjunx 1	30.07				
musdjunx 721	27.84				
tgtgcactaaaatggaacagcccttctac					
2 29	1 1 2	2 2 2 2 3 4			
humbjunx 65533	60.68				
musbjunx 65533	49.58				
muscjunx 1	29.97				
musdjunx 721	27.66				
gtgcactaaaatggaacagcccttctac					
3 28	1 1 2	2 2 2 2 3 4			
humbjunx 65533	60.60				
musbjunx 65533	49.10				
muscjunx 1	29.86				
musdjunx 721	27.47				
tgactaaaatggaacagcccttctacc					
4 28	1 1 1 1	2 2 2 2 3 4			
humbjunx 65533	60.60				
musbjunx 65533	46.57				
muscjunx 1	29.86				
musdjunx 729	27.47				
gcactaaaatggaacagcccttctacc					
5 27	1 1 1 1	2 2 2 2 3 4			
humbjunx 5	60.51				
musbjunx 5	45.96				
muscjunx 1	29.75				
musdjunx 729	27.26				

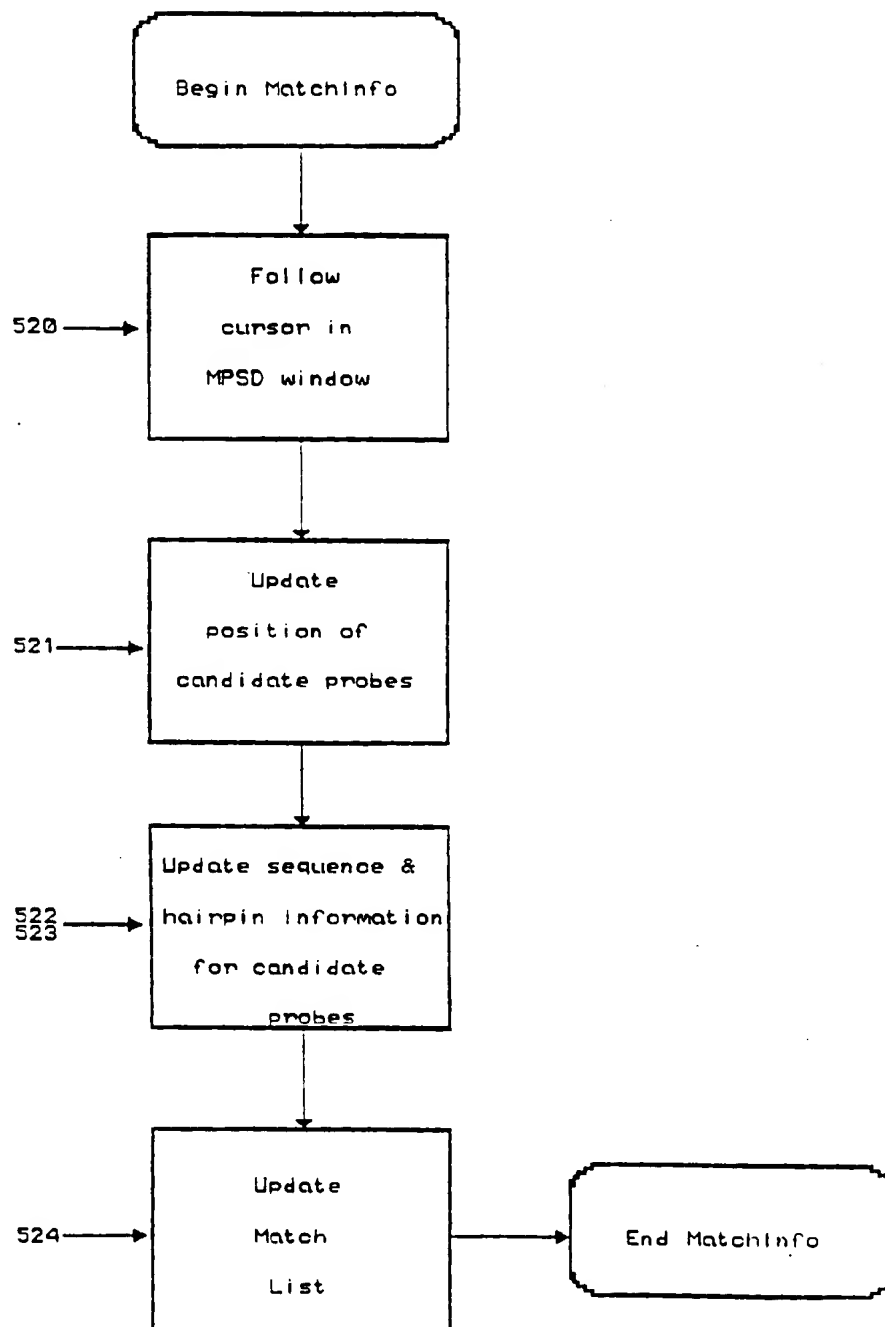
54/59

FIG. 45



55/59

FIG. 46



56/59

FIG. 47

LOCUS	HUMBJUNX	1044 bp	DNA		19-DEC-1991	
BASE COUNT	195 A	368 C	340 G	141 T		
ORIGIN						
1	ATGTGCACTA	AAATGGAACA	GCCCTTCTAC	CACGACGACT	CATACACAGC	TACGGGATAC
61	GGCCGGGCCC	CTGGTGGCCT	CTCTCTACAC	GA CTACAAAC	TCCTGAAACC	GAGCCTGGCG
121	GTCAACCTGG	CCGACCCCTA	CCGGAGTCTC	AAAGCGCCTG	GGGCTCGCGG	ACCCGGCCCA
181	GAGGGCGGCG	GTGGCGGCAG	CTACTTTTCT	GGTCAGGGCT	CGGACACCGG	CGCGTCTCTC
241	AAGCTCGCCT	CTTCGGAGCT	GGAACGCCTG	ATTGTCCCCA	ACAGCAACGG	CGTGATCACG
301	ACGACGCCTA	CACCCCCGGG	ACAGTACTTT	TACCCCCGCG	GGGGTGGCAG	CGGTGGAGGT
361	GCAGGGGGCG	CAGGGGGCGG	CGTCACCGAG	GAGCAGGAGG	GCTTCGCCGA	CGGCTTTGTC
421	AAAGCCCTGG	ACGATCTGCA	CAAGATGAAC	CACGTGACAC	CCCCCAACGT	GTCCCTGGGC
481	GCTACCGGGG	GGCCCCCGGC	TGGGCCCCGG	GGCGTCTACG	CCGGCCCCGA	GCCACCTCCC
541	GTTTACACCA	ACCTCAGCAG	CTACTCCCCA	GCCTCTGCGT	CCTCGGGAGG	CGCCGGGGCT
601	GCCGTCGGGA	CCGGGAGCTC	GTACCCGACG	ACCACCATCA	GCTACCTCCC	ACACGCGCCG
661	CCCTTCGCGG	GTGGCCACCC	GGCGCAGCTG	GGCTTGGGCC	GCGGCGCCTC	CACCTTCAAG
721	GAGGAACCGC	AGACCGTGCC	GGAGGCGCGC	AGCCGGGACG	CCACGCCGCC	GGTGTCCTCC
781	ATCAACATGG	AAGACCAAGA	GCGCATCAAA	GTGGAGCGCA	AGCGGCTGCG	GAACCGGCTG
841	GCGGCCACCA	AGTGCCGGAA	GCGGAAGCTG	GAGCGCATCG	CGCGCCTGGA	GGACAAGGTG
901	AAGACGCTCA	AGGCCGAGAA	CGCGGGGCTG	TCGAGTACCG	CCGGCCTCCT	CCGGGAGCAG
961	GTGGCCCAGC	TCAAACAGAA	GGTCATGACC	CACGTCAGCA	ACGGCTGTCA	GCTGCTGCTT
1021	GGGGTCAAGG	GACACGCCTT	CTGA			

//

57/59

FIG. 48

LOCUS HUMBJUNX 1044 bp DNA 19-DEC-1991  
 BASE COUNT 195 A 368 C 340 G 141 T  
 ORIGIN

1 ATGTGCACTA AAATGGAACA GCCCTTCTAC CACGACGACT CATACACAGC TACGGGATAC  
 61 GGCCGGGGCC CTGGTGGCCT CTCTCTACAC GACTACAAAC TCCTGAAACC GAGCCTGGCG  
 121 GTCAACCTGG CCGACCCCTA CCGGAGTCTC AAAGCGCCTG GGGCTCGCGG ACCCGGCCCA  
 181 GAGGGCGGCG GTGGCGGCAG CTACTTTTCT GGTCAGGGCT CGGACACCGG CGCGTCTCTC  
 241 AAGCTCGCCT CTTCGGAGCT GGAACGCCTG ATTGTCCCCA ACAGCAACGG CGTGATCACC  
 301 ACGACGCCTA CACCCCGGGG ACAGTACTTT TACCCCGCGG GGGTGGCAG CGGTGGAGGT  
 361 GCAGGGGGCG CAGGGGGCGG CGTCACCGAG GAGCAGGAGG GCTTCGCCGA CGGCTTTGTC  
 421 AAAGCCCTGG ACGATCTGCA CAAGATGAAC CACGTGACAC CCCCCAACGT GTCCCTGGGG  
 481 GCTACCGGGG GGCCCCCGG TGGGCCCGGG GGCGTCTACG CCGGCCCGGA GCCACCTCCC  
 541 GTTTACACCA ACCTCAGCAG CTACTCCCCA GCCTCTGCGT CCTCGGGAGG CGCCGGGGCT  
 601 GCCGTCGGGA CCGGGAGCTC GTACCCGACG ACCACCATCA GCTACCTCCC ACACGCGCCG  
 661 CCCTTCGCCG GTGGCCACCC GGCGCAGCTG GGCTTGGGCC GCGGCGCCTC CACCTTCAAG  
 721 GAGGAACCGC AGACCGTGCC GGAGGCGCGC AGCCGGGACG CCACGCGCGG GGTGTCCCCC  
 781 ATCAACATGG AAGACCAAGA GCGCATCAA GTGGAGCGCA AGCGGCTGCG GAACCGGCTG  
 841 GCGGCCACCA AGTGCCGGAA GCGGAAGCTG GAGCGCATCG CGCGCTGGA GGACAAGGTG  
 901 AAGACGCTCA AGGCCGAGAA CGCGGGGCTG TCGAGTACCG CCGGCTCCTT CCGGGAGCAG  
 961 GTGGCCACAG TCAAACAGAA GGTTCATGAC CACGTCAGCA ACGGCTGTCA GCTGCTGCTT  
 1021 GGGGTCAAGG GACACGCCTT CTGA

/

LOCUS HUMCJUNX 996 bp DNA 19-DEC-1991  
 BASE COUNT 226 A 342 C 299 G 129 T  
 ORIGIN

1 ATGACTGCAA AGATGGAAC GACCTTCTAT GACGATGCCC TCAACGCCTC GTTCCTCCCG  
 61 TCCGAGAGCG GACCTTATGG CTACAGTAAC CCCAAGATCC TGAAACAGAG CATGACCCTG  
 121 AACCTGGCCG ACCCAGTGGG GAGCCTGAAG CCGCACCTCC GCGCCAAGAA CTCGGACCTC  
 181 CTCACCTCGC CCGACGTGGG GCTGCTCAAG CTGGCGTCGC CCGAGCTGTA GCGCCTGATA  
 241 ATCCAGTCCA GCAACGGGCA CATCACCACC ACGCCGACCC CCACCCAGTT CCTGTGCCCC  
 301 AAGAACGTGA CAGATGAGCA GGAGGGGTTT CCGGAGGGCT TCGTGCCTGC CCTGGCCGAA  
 361 CTGCACAGCC AGAACACGCT GCCCAGCGTC ACGTCGGCGG CGCAGCCGGT CAACGGGGCA  
 421 GGCATGGTGG CTCCCGCGGT AGCCTCGGTG GCAGGGGGCA GCGGCAGCGG CCGCTTCAGC  
 481 GCCAGCCTGC ACAGCGAGCC GCCGGTCTAC GCAAACCTCA GCAACTTCAA CCCAGGCGCG  
 541 CTGAGCAGCG GCGGCGGGGC GCCCTCCTAC GGCGCGGCGG GCCTGGCCTT TCCCGCGCAA  
 601 CCCCAGCAGC AGCAGCAGCC GCCGCACCAC CTGCCCCAGC AGATGCCCCG GCAGCACCAG  
 661 CGGCTGCAGG CCCTGAAGGA GGAGCCTCAG ACAGTGCCCG AGATGCCCCG CGAGACACCG  
 721 CCCCTGTCCC CCATCGACAT GGAGTCCCAG GAGCGGATCA AGGCGGAGAG GAAGCGCATG  
 781 AGGAACCGCA TCGCTGCCTC CAAGTGCCGA AAAAGGAAGC TGGAGAGAAT CGCCCGGCTG  
 841 GAGGAAAAAG TGAAAACCTT GAAAGCTCAG AACTCGGAGC TGGCGTCCAC GGCCAACATG  
 901 CTCAGGGAAC AGGTGGCACA GCTTAAACAG AAAGTCATGA ACCACGTTAA CAGTGGGTGC  
 961 CAACTCATGC TAACGCAGCA GTTGCAAACA TTTTGA

LOCUS HUMDJUNX 1044 bp ss-mRNA PRI 24-MAY-1991  
 DEFINITION Human junD mRNA  
 X56681  
 KEYWORDS jun-D gene; oncogene.  
 SOURCE Homo sapiens RNA.  
 ORGANISM Homo sapiens  
 Eukaryota; Animalia; Metazoa; Chordata; Vertebrata; Mammalia;  
 Theria; Eutheria; Primates; Haplorhini; Catarrhini; Hominidae.  
 REFERENCE 1 (bases 1 to 1891)  
 AUTHORS Shaul, Y.  
 JOURNAL Unpublished (1990)  
 STANDARD full automatic  
 REFERENCE 2 (sites)

58/59

FIG. 48 (2)

AUTHORS Berger, I. and Shaul, Y.  
 TITLE Structure and function of human jun-D  
 JOURNAL Unpublished (1990)  
 STANDARD full staff review  
 COMMENT From EMBL 26 entry H5JUNDR; dated 18-MAR-1991.

FEATURES Location/Qualifiers  
 mRNA 1..1891  
 /gene="junD"  
 /evidence=EXPERIMENTAL  
 CDS 175..1218  
 /product="junD protein"  
 /gene="junD"  
 /codon\_start=1  
 polyA\_site 1891..1891  
 BASE COUNT 162 A 405 C 360 G 117 T  
 ORIGIN

```

1 ATGGAAACAC CCTTCTACGG CGATGAGGCG CTGAGCGGCC TGGGCGGCGG CGCCAGTGGC
61 AGCGGCGGCA CGTTCGCGTC CCGGGGCGGC TTGTTCCCCG GGGCGCCCCC GACGGCCGCG
121 GCCGGCAGCA TGATGAAGAA GGACGCGCTG ACGCTGAGCC TGAGTGAGCA GGTGGCGGCA
181 GCGCTCAAGC CTGCGCCCGC GCCCGCCTCC TACCCCCCTG CCGCCGACGG CGCCCCCAGC
241 GCGGCACCCC CCCACGGCCT GCTCGCCTCT CCCGACCTGG GGCTGCTGAA GCTGGCCTCC
301 CCGGAGCTCG AGCGCCTCAT CATCCAGTCC AACGGGCTGG TCACCACCAC GCCGACGAGC
361 TCACAGTTCC TCTACCCCAA GGTGGCGGCC AGCGAGGAGC AGGAGTTCCG CGAGGGCTTC
421 GTCAAGGCCC TGGAGGATTT ACACAAGCAG AACCAGCTCG GCGCGGGCCG GGCCGCTGCC
481 GCCGCCGCGG CCGCGCGCGG GGGGCCCTCG GGCACGGCCA CGGGCTCCGC GCGCCCCGGC
541 GAGCTGGCCC CCGCGGGCGG CGCGCCCGAA GCGCCTGTCT ACGCGAACCT GAGCAGCTAC
601 GCGGGCGGCG CCGGGGGCGC GGGGGGGCGC GCGACGGTCG CCTTCGCTGC CGAACCTGTG
661 CCCTTCCCGC CGCCGCCACC CCCAGGCGCG TTGGGGCCGC CGCGCCTGGC TCGGCTCAAG
721 GACGAGCCAC AGACGGTGCC CGACGTGCCG AGCTTCGGCG AGAGCCCGCC GTTGTGCCCC
781 ATCGACATGG ACACGCAGGA GCGCATCAAG GCGGAGCGCA AGCGGCTGCG CAACCGCATC
841 GCCGCCCTCA AGTGCCGCAA GCGCAAGCTG GAGCGCATCT CGCGCCTGGA AGAGAAAGTG
901 AAGACCCTCA AGAGTCAGAA CACGAGCTG GCGTCCACGG CGAGCCTGCT GCGCGAGCAG
961 GTGGCGCAGC TCAAGCAGAA AGTCCTCAGC CACGTCAACA CCGGCTGCCA GCTGCTGCCC
1021 CAGCACCAGG TCCCGGCGTA CTGA

```

LOCUS MUSBJUNX 1035 bp DNA 19-DEC-1991  
 BASE COUNT 210 A 333 C 333 G 159 T  
 ORIGIN

```

1 ATGTGCACGA AAATGGAACA GCCTTTCTAT CACGACGACT CTTACGCAGC GGCGGGATAC
61 GGTCGGAGCC CTGGCAGCCT GTCTCTACAC GACTACAAAC TCCTGAAACC CACCTTGGCG
121 CTCAACCTGG CGGATCCCTA TCGGGGTCTC AAGGGTCTTG GGGCGCGGGG TCCAGGCCCG
181 GAGGGCAGTG GGGCAGGCAG CTACTTTTCG GGTCAGGGAT CAGACACAGG CGCATCTCTG
241 AAGCTAGCCT CCACGGA ACT GGAGCGCTTG ATCGTCCCCA ACAGCAACGG CGTGATCAGC
301 ACGACGCCCC CGCCTCCGGG ACAGTACTTT TACCCCGTGG GGGGTGGCAG CGGTGGAGGT
361 ACAGGGGGCG GCGTCACCGA GGAGCAGGAG GGCTTTGCGG ACGGTTTGTG CAAAGCCCTG
421 GACGACCTGC ACAAGATGAA CCACGTGACG CCCCCCAACG TGTCCTGGGG CGCCAGCGGG
481 GGTCCCCAGG CCGGCCCAGG GGGCGTCTAT GCTGGTCCGG AGCCGCTCC CGTCTACACC
541 AACCTCAGCA GTTACTCTCC AGCCTCTGCA CCCTCTGGAG GCTCCGGGAG CGCCGTCGGG
601 ACTGGGAGCT CATACCCGAC GGGCACCATC AGCTACCTCC CACATGCACC ACCCTTTGCG
661 GGCGGCCACC CGGCACAGCT GGGTTTGAGT CGTGGCGCTT CCGCCTTTAA AGAGGAACCG
721 CAGACCGTAC CCGAGGCACG CAGCCGCGAC GCCACGCGCG CTGTGTCCCC CATCAACATG
781 GAAGACCAGG AGCGCATCAA AGTGGAGCGA AAGCGGCTGC GGAACAGGCT GGCGGCCACC
841 AAGTGCCGGA AGCGGAAGCT GGAGCGCATC GCGCGCCTGG AGGACAAGGT GAAGACATC
901 AAGGCTGAGA ACGCGGGGCT GTCGAGTGCT GCCGGTCTCC TAAGGGAGCA AGTGGCGCAG
961 CTCAAGCAGA AGGTCATGAC CCATGTCAGC AACGGCTGCC AGTTGCTGCT AGGGGTCAAG
1021 GGACACGCTT TCTGA

```



FIG. 48 (3)

OCUS MUSCJUNX 1005 bp DNA 19-DEC-1991  
 ASE COUNT 223 A 334 C 300 G 148 T  
 RIGIN

```

1 ATGACTGCAA AGATGGAAAC GACCTTCTAC GACGATGCCC TCAACGCCCTC GTTCCTCCAG
61 TCCGAGAGCG GTGCCCTACGG CTACAGTAAC CCTAAGATCC TAAAACAGAG CATGACCTTG
121 AACCTGGCCG ACCCGGTGGG CAGTCTGAAG CCGCACCTCC GCGCCAAGAA CTCGGACCTT
181 CTCACGTCCG CCGACGTCCG GCTGCTCAAG CTGGCGTCCG CGGAGCTGGA GCGCCTGATC
241 ATCCAGTCCA GCAATGGGCA CATCACCCT ACACCGACCC CCACCCAGTT CTTGTGCCCCC
301 AAGAACGTGA CCGACGAGCA GGAGGGCTTC GCCGAGGGCT TCGTGCCTGC CCTGGCTGAA
361 CTGCATAGCC AGAACACGCT TCCCAGTGTC ACCTCCGCGG CACAGCCGCT CAGCGGGGCG
421 GGCATGGTGG CTCCCGCGGT GGCCTCAGTA GCAGGCGCTG GCGGCGGTGG TGGCTACAGC
481 GCCAGCCTGC ACAGTGAGCC TCCGGTCTAC GCCAACCTCA GCAACTTCAA CCGGGGTGCG
541 CTGAGCAGCG GCGGTGGGGC GCCCTCCTAT GGCGCGGCGG GGCTGGCCTT TCCCTCCGAG
601 CCGCAGCAGC AGCAGCAGCC GCCTCAGCCG CCGCACCACT TGCCCCAACA GATCCCGGTG
661 CAGCACCCGC GGCTGCAAGC CCTGAAGGAA GAGCCGCAGA CCGTGGCCGA GATCCCGGGA
721 GAGACGCCGC CCCTGTCCCC TATCGACATG GAGTCTCAGG AGCGGATCAA GGCAGAGAGG
781 AAGCGCATGA GGAACCGCAT TGCCGCTTCC AAGTGCCTGA AAAGGAAGCT GGAGCGGATC
841 GCTCGGCTAG AGGAAAAAGT GAAACCTTG AAAGCGCAA ACTCCGAGCT GGCATCCACG
901 GCCAACATGC TCAGGGAACA GGTGGCAGAG CTTAAGCAGA AAGTCATGAA CCACGTAAAC
961 AGTGGGTGCC AACTCATGCT AACGCAGCAG TTGCAAACGT TTGGA
  
```

MUSCJUNX 1026 bp DNA 19-DEC-1991  
 ASE COUNT 172 A 382 C 343 G 129 T  
 RIGIN

```

1 ATGGAACGCG CCTTCTATGG CGAGGAGGCG CTGAGCGGCC TGGCTGCGGG TGGCTCGAGC
61 GTCGCTGGTG CTAAGTGGGG CCGCGGCGGT GGTGGCTTCG CGCCCCCGGG CCGCGCTTTC
121 CCGGGGGCGC CCGGACCGAG CAGCATGCTG AAGAAAGACG CGCTGACGCT CAGCCTGGCG
181 GAGCAGGGAG CGGCGGGATT GAAACCAGGG TCGGCCACTG CACCTTCTGC GCTGCGCCCCC
241 GACGGCGCCC CCGACGGGCT GCTGGCTTCG CCGGATCTTG GGCTGCTCAA ACTCGGCTCG
301 CCGGAGCTGG AGAGGCTGAT CATCCAGTCC AACGGGCTGG TGACCACTAC CCGGACCACT
361 ACGCAGTTCC TCTACCCGAA GGTGGCAGCC AGCGAGGAGC AGGAGTTCCG CGAAGGCTTC
421 GTCAAGGCGC TGGAGGACCT GCACAAGCAA AGCCAGCTGG GTGCGGCCAC CGCGGCCACC
481 TCAGGGGCTC CCGCGCCTCC CCGCGCCGCC GACCTGGCCG CCACCCCGGG GGCCACGGAG
541 ACCCGGCTCT ACGCCAACCT GAGCAGTTTC GCGGGTGGCG CCGGGCCCCC TGGGGGCGCG
601 GCCACCGTGG CTTTCGCGCG GGAGCCAGTG CCCTTCCCGC CGCCCCCGGG CGCGCTGGGG
661 CCGCGGCCAC CTCCGCATCC ACCGCGCCTG GCGCGCTCA AGGACGAGCC GCAGACCGTG
721 CCGGACGTGC CGAGCTTCGG CGACAGCCCT CCGCTGTCCG CCATCGACAT GGACACGCAA
781 GAACGCATCA AGGCGGAGCG CAAGAGGCTG CGCAACCGCA TCGCCGCTC CAAATGCCGC
841 AAGCGCAAGC TGGAGCGTAT CTCGCGCCTG GAGGAGAAAG TCAAGACCTT CAAAAGCCAG
901 AACACCGAGC TGGCGTCCAC CGCCAGCCTG CTGCGCGAGC AGGTGGCGCA GCTCAAACAG
961 AAAGTCCTCA GCCACGTCAA CAGCGGCTGC CAGCTGCTGC CCCAGCACC GGTCCCGGGC
1021 TACTGA
  
```

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US93/00977

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(5) : C12Q 1/44, 1/68; C07H 21/00

US CL : 435/6, 19; 536/24.3

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6, 19; 536/24.3

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Analytical Biochemistry, Volume 169, issued 1988, J. A. Matthews et al, "Review: Analytical Strategies for the Use of DNA Probes", pages 1-25, see entire document.	1-28, 33-40
Y	Proceedings of the National Academy of Science, USA, Volume 84, issued October 1987, L. J. Emorine et al., "Structure of the gene for human $\beta_2$ -adrenergic receptor: Expression and promoter characterization", pages 6995-6999, especially Figure 1.	9, 33, 34
Y	US, A, 4,729,950 (Kricka et al.) 08 March 1988, see entire document.	1-28, 33-40



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:	*T	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
*A* document defining the general state of the art which is not considered to be part of particular relevance	*X*	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
*E* earlier document published on or after the international filing date	*Y*	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
*L* document which may throw doubt on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*G*	document member of the same patent family
*O* document referring to an oral disclosure, use, exhibition or other means		
*P* document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search

04 May 1993

Date of mailing of the international search report

13 MAY 1993

Name and mailing address of the ISA/US  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231

Facsimile No. NOT APPLICABLE

Authorized officer

DAVID SCHREIBER

Telephone No. (703) 308-0196

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US93/00977

## C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Gene Analysis Techniques, Volume 4, issued 1987, C. B. Harley, "Hybridization of Oligo(dT) to RNA on Nitrocellulose", pages 17-22, see entire document.	1-28, 33-40
X Y	Methods in Enzymology, Vol. 65, Pt. 1, issued 1980, A. R. Dunn et al, "Mapping Viral mRNAs by Sandwich Hybridization", pages 468-478, especially page 46, second full paragraph.	<u>1-3, 6</u> 4, 5, 7-28, 33-40
X Y	DNA, Volume 7, No. 2, issued 1988, A. Palva et al, "Laboratory Methods: Quantification of $\alpha$ -Amylase mRNA in <i>Bacillus subtilis</i> by Nucleic Acid Sandwich Hybridization", pages 135-142, see entire document.	<u>1, 2, 6</u> 4, 5, 7-28, 33-40
X Y	US, A, 4,957,865 (Samarut et al) 18 September 1990, col. 11, lines 34-41.	<u>29-31</u> 32
Y	Journal of Molecular Evolution, Volume 19, Numbers 3-4, R. Nussinov, "Efficient algorithms for searching for exact repetition of nucleotide sequences.", pages 283-285, see Abstract 05061573.	23
Y	Computer Applications in the Biosciences, Volume 8, No. 3, issued September 1987, W. W. Ralph et al, "A modified Chou and Fasman protein structure algorithm", pages 211-216, see Abstract 06625157.	23-25
Y	Biochemical and Biophysical Research Communications, Volume 170, No. 1, issued 1990, C. Hough et al, "Differential down-regulation of $\beta_1$ - and $\beta_2$ -adrenergic receptor mRNA in C <sub>6</sub> glioma cells", pages 46-52, see entire document.	9, 33, 34
Y	Analytical Biochemistry, Volume 196, issued 1991, A. Ballagi-Pordany et al, "Quantitative determination of mRNA phenotypes by the polymerase chain reaction", pages 89-94, see especially the Abstract.	1-3
Y	Journal of Biological Chemistry, Volume 265, No. 27, issued 25 September 1990, B. Feve et al, "Differential regulation of $\beta_1$ - and $\beta_2$ -adrenergic receptor protein and mRNA levels by glucocorticoids during 3T3-F442A adipose differentiation", pages 16343-16349, see entire document.	9, 33, 34
Y	Journal of Biological Chemistry, Volume 266, No. 35, issued 15 December 1991, P. Muzzin et al, "An adipose tissue-specific $\beta$ -adrenergic receptor", pages 24053-24058, especially Figures 1 and 2.	9, 33, 34

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US93/00977

## C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Analytical Biochemistry, Volume 170, issued 1988, C. Albretsen et al, "Optimal conditions for hybridization with oligonucleotides: A study with <i>myc</i> -oncogene DNA probes", pages 193-202, especially paragraph bridging pages 193 and 194.	1-28
Y	Proceedings of the National Academy of Science USA, Volume 84, issued March 1987, M. S. Waterman et al, "Phase transitions in sequence matches and nucleic acid structure", pages 1239-1243, see entire document.	1-28
Y	Journal of Molecular Biology, Volume 197, issued 1987, M. S. Waterman et al, "A new algorithm for best subsequence alignments with application to tRNA-rRNA comparisons", pages 723-728, see entire document.	1-23
Y	Biotechniques, Volume 12, No. 2, issued Feb 1992, R. J. Cano et al, "DNA hybridization assay using ATTOPHOS <sup>TM</sup> , a fluorescent substrate for alkaline phosphatase", pages 264-269, see entire document.	1, 2, 18, 21, 22
Y	Genomics, Vol. 11, issued 1991, W. R. Pearson, "Searching protein sequence libraries: Comparison of the sensitivity and selectivity of the Smith-Waterman and FASTA algorithms", pages 635-650, see entire document.	1-23
Y	Biochemistry, Vol. 30, issued 1991, N. P. Gerard, "Human substance P receptor (NK-1): Organization of the gene, chromosome localization, and functional expression of cDNA clones", pages 10640-10646, especially Figure 1.	12, 39, 40
Y	The EMBO Journal, Volume 8, No. 5, issued 1989, S.-I. Hirai et al, "Characterization of <i>junD</i> : a new member of the <i>jun</i> proto-oncogene family", pages 1433-1439, especially Figure 1b.	35, 36
Y	Journal of Biological Chemistry, Volume 262, No. 29, issued 15 October 1987, D. T. Jones et al, "Molecular cloning of five GTP-binding protein cDNA species from rat olfactory neuroepithelium", pages 14241-14249, especially Figures 2-6.	10, 11, 13-15, 37-38
Y	Journal of Biological Chemistry, Volume 265, No. 3, issued 25 January 1990, L. Li et al, "Different members of the <i>jun</i> proto-oncogene family exhibit distinct patterns of expression in response to type $\beta$ transforming growth factor", pages 1556-1562, especially Figure 5.	35, 36

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US93/00977

## C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Journal of Biological Chemistry, Volume 264, No. 30, issued 25 October 1989, Y. Yokota et al, "Molecular characterization of a functional cDNA for rat substance P receptor", pages 17649-17652, especially Figure 2.	12, 39, 40
Y	Journal of Biological Chemistry, Volume 263, No. 14, issued 15 May 1988, J. Codina et al, " $\alpha_i$ -3 cDNA encodes the $\alpha$ subunit of $G_k$ , the stimulatory G protein of receptor-regulated $K^+$ channels", pages 6746-6750, especially Figure 2.	13, 37, 38
Y	Journal of Biological Chemistry, Volume 263, No. 14, issued 15 May 1988, H. Itoh et al, "Presence of three distinct molecular species of $G_i$ protein $\alpha$ subunit", pages 6656-6664, especially Figures 2, 4-6, and 8.	10, 11, 13-15, 37, 38
Y	Proceedings of the National Academy of Science USA, Volume 85, issued December 1988, K. Hattori et al, "Structure and chromosomal localization of the functional intronless human <u>JUN</u> protooncogene", pages 9148-9152, especially Figure 5.	35, 36
Y	Proceedings of the National Academy of Science USA, Volume 84, issued August 1987, P. Bray et al, "Human cDNA clones for an $\alpha$ subunit of $G_i$ signal-transduction protein", pages 5115-5119, especially Figure 2 and 5-7.	10, 11, 13-15, 37, 38
Y	Proceedings of the National Academy of Science USA, Volume 84, issued November 1987, C. R. Beals et al, "A small multigene family encodes $G_i$ signal-transduction proteins", pages 7886-7890, especially Figures 1, 2, and 4.	10, 11, 13-15, 37, 38
Y	Nucleic Acids Research, Volume 16, No. 8, issued 1988, B. A. Harris, "Complete cDNA sequence of a human stimulatory GTP-binding protein alpha subunit", page 3585.	10, 11, 13-15, 37, 38
Y	FEBS Letters, Volume 206, No. 1, issued September 1986, R. Mattera et al, "Identification by molecular cloning of two forms of the $\alpha$ -subunit of the human liver stimulatory ( $G_i$ ) regulatory component of adenylyl cyclase", pages 36-42, especially Figure 2.	10, 11, 13-15, 37, 38
Y	FEBS Letters, Volume 219, No. 1, issued July 1987, J. R. Didsbury et al, "Molecular cloning of a new human G protein", pages 259-263, especially Figures 1 and 2.	10, 11, 13-15, 37, 38
Y	Nucleic Acids Research, Volume 19, No. 17, issued 1991, A. Swaroop et al, "Differential expression of novel $G_{s\alpha}$ signal transduction protein cDNA species", pages 4725-4729, see entire document.	10, 11, 13-15, 37, 38

**B. FIELDS SEARCHED**

Electronic data bases consulted (Name of data base and where practicable terms used):

N-GENESEQ, GENBANK, GENBANK-NEW, UEMBL, EMBL-NEW, CA, MEDLINE, BIOSIS, APS, search terms: Waterman, Pevzner, second hybridization strength, continuous seed filtration, mitsubishi probe selection, calculation of melting temperature of nucleic acids, H-site model, nucleation threshold, mismatch model, measuring or assaying beta receptor mRNA, measuring or assaying substance P mRNA, measuring or assaying jun mRNA, measuring or assaying Gs, Go, or Gi protein mRNA

**BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING**

This ISA found multiple inventions as follows:

- I. Claims 1-9, 16-28, 33 and 34, drawn to a method for detecting and quantifying specific mRNAs and a polynucleotide immobilized support for use in said method which detects  $\beta$ -receptor mRNA, classified in Class 435/6, for example.
- II. Claim 10, drawn to a method for detecting and quantifying specific Gs protein mRNA, classified in Class 435/6, for example.
- III. Claim 11, drawn to a method for detecting and quantifying specific Gi-2 protein mRNA, classified in Class 435/6, for example.
- IV. Claims 12, 39, and 40, drawn to a method for detecting and quantifying specific substance P mRNA and polynucleotide support for accomplishing the method, classified in Class 435/6, for example.
- V. Claim 13, drawn to a method for detecting and quantifying specific Gi-3 protein mRNA, classified in Class 435/6, for example.
- VI. Claim 14, drawn to a method for detecting and quantifying specific Gi-1 protein mRNA, classified in Class 435/6, for example.
- VII. Claims 15, 37, and 38, drawn to a method for detecting and quantifying specific G protein mRNA and polynucleotide immobilized support for accomplishing the method, classified in Class 435/6 for example.
- VIII. Claims 35 and 36, drawn to a polynucleotide immobilized support for the detection and quantification of jun-oncogene mRNA, classified in Class 536/24.3.
- IX. Claims 29-32, drawn to a method of inhibiting RNase activity in a sample, classified in Class 435/19.